THE UNIVERSITY *of* EDINBURGH

On the Performance of Hybrid Beamforming for Millimeter Wave Wireless Networks

Oluwatayo Yetunde Kolawole

A thesis submitted for the degree of Doctor of Philosophy

Institute for Digital Communications

University of Edinburgh

March, 2019

Here I raise my Ebenezer

Hither by Thy Help I've come!

–Robert Robinsin–

For my Father, who taught me to dream

Declaration of Originality

I hereby declare that the research recorded in this thesis and the thesis itself were composed and originated entirely by myself in the Institute for Digital Communications of the School of Engineering at The University of Edinburgh.

Oluwatayo Yetunde Kolawole

Edinburgh, UK

March, 2019

Acknowledgements

First, I would like to express my deepest appreciation to my supervisor Prof. Tharmalingam Ratnarajah for being a constant support to me throughout my studies. I am grateful for his enthusiasm and priceless suggestions, which have made this work so much richer.

I would also like to acknowledge the former president of Nigeria, Dr Goodluck Jonathan, for initiating a scheme that enabled me to secure the necessary funding for this research. In addition, the petroleum technology development fund has provided the generous financial support throughout this PhD programme.

Thanks to all my colleagues at the Institute for Digital Communications, it has been wonderful meeting and conversing with many great minds. Thanks to members of my group, Satyanarayana Vuppala, Anastasios Papazafeiropoulos, Sudip Biswas, Keshav Singh for their technical advice and input.

Last but not least, this PhD could not have been completed without the love and support of my family and friends. My heartfelt gratitude goes to my mum, Dr. Susan Kolawole, and my siblings, Oluwabukola Olaniyan, Babatunde Kolawole, Oluwabunmi Adeojo, Oloruntobiloba Kolawole and their respective families. Your love, encouragement and prayers are what put me through school. To my friends, Ella Edwin, Dayo Mofikoya, Furaha Asani, Helen Cope, Dindul Dadi, Ania Ogenyi, Chidie and Charlene Okwudire, Janet Marriott, you've provided a listening ear, encouraged me and prayed for me. I'm grateful for your friendship. To David and Dorothy Cain, thank you for opening up your home to me, I can't express how refreshing the weekends spent with you were. To my small group, you've become family to me. Thank you for coming alongside me on this journey. To my husband, Ayomide Adebayo, thank you for moving across continents, and putting up with my late nights. Your understanding and kindness are invaluable to me.

Abstract

The phenomenal growth in the demand for mobile wireless data services is pushing the boundaries of modern communication networks. Developing new technologies that can provide unprecedented data rates to support the pervasive and exponentially increasing demand is therefore of prime importance in wireless communications. In existing communication systems, physical layer techniques are commonly used to improve capacity. Nevertheless, the limited available resources in the spectrum are unable to scale up, fundamentally restricting further capacity increase. Consequently, alternative approaches which exploit both unused and underutilised spectrum bands are highly attractive. This thesis investigates the use of the millimeter wave (mmWave) spectrum as it has the potential to provide unlimited bandwidth to wireless communication systems.

As a first step toward realising mmWave wireless communications, a cloud radio access network using mmWave technology in the fronthaul and access links is proposed to establish a feasible architecture for deploying mmWave systems with hybrid beamforming. Within the context of a multi-user communication system, an analytical framework of the downlink transmission is presented, providing insights on how to navigate across the challenges associated with high-frequency transmissions. The performance of each user is measured by deriving outage probability, average latency and throughput in both noise-limited and interference-limited scenarios. Further analysis of the system is carried out for two possible user association configurations. By relying on large antenna array deployment in highly dense networks, this architecture is able to achieve reduced outages with very low latencies, making it ideal to support a growing number of users.

The second part of this work describes a novel two-stage optimisation algorithm for obtaining hybrid precoders and combiners that maximise the energy efficiency (EE) of a general multi-user mmWave multiple-input, multiple-output (MIMO) interference channel network involving internet of things (IoT) devices. The hybrid transceiver design problem considers both perfect and imperfect channel state information (CSI). In the first stage, the original non-convex multivariate EE maximization problem is transformed into an equivalent univariate problem and the optimal single beamformers are then obtained by exploiting the correlation between parametric and fractional programming problems and the relationship between weighted sum rate (WSR) and weighted minimum mean squared error (WMMSE) problems. The second stage involves the use of an orthogonal matching pursuit (OMP)-based algorithm to obtain the energy-efficient hybrid

beamformers. This approach produces results comparable to the optimal beamforming strategy but with much lower complexity, and further validates the use of mmWave networks in practice to support the demand from ubiquitous power-constrained smart devices.

In the third part, the focus is on the more practical scenario of imperfect CSI for multi-user mmWave systems. Following the success of hybrid beamforming for mmWave wireless communication, a non-traditional transmission strategy called Rate Splitting (RS) is investigated in conjunction with hybrid beamforming to tackle the residual multi-user interference (MUI) caused by errors in the estimated channel. Using this technique, the transmitted signal is split into a common message and a private message with the transmitted power dynamically divided between the two parts to ensure that there is interference-free transmission of the common message. An alternating maximisation algorithm is proposed to obtain the optimal common precoder. Simulation results show that the RS transmission scheme is beneficial to multi-user mmWave transmissions as it enables remarkable rate gains over the traditional linear transmission methods.

Finally, the fourth part analyses the spectral efficiency (SE) performance of a mmWave system with hybrid beamforming whilst accounting for real-life practice transceiver hardware impairments. An investigation is conducted into three major hardware impairments, namely, the multiplicative phase noise (PN), the amplified thermal noise (ATN) and the residual additive transceiver hardware impairments (RATHI). The hybrid precoder is designed to maximise the SE by the minimisation of the Euclidean distance between the optimal digital precoder and the noisy product of the hybrid precoders while the hybrid combiners are designed by the minimisation of the mean square error (MSE) between the transmitted and received signals. Multiplicative PN was found to be the most critical of the three impairments considered. It was observed that the additive impairments could be neglected for low signal-to-noise-ratio (SNR) while the ATNs caused a steady degradation to the SE performance.

Lay Summary

Mobile communications have transformed the world. Over the last decade, with the growing wave of wireless data services (such as social media, multimedia streaming and virtual reality), it has become an expectation that vast amounts of data be available in all places at all times with increasing speeds. This pervasive demand for mobile data is stretching the limits of existing communication systems whilst driving the need for the development of new communication systems with the ability to respond to vast data traffic and a growing number of connected devices. Against this backdrop, the primary goals of the fifth-generation (5G) mobile communications relative to the existing fourth-generation (4G) communication networks are to provide connections for a minimum of 100 billion devices, a 1000-fold increase capacity and very high data rates for individual users (10 gigabits per second). To achieve these enormous goals, new technologies are required.

This thesis addresses this issue by examining millimeter wave (mmWave) communications, a paradigm-shifting technology, which operate on the high-frequency spectrum bands not yet used for cellular communications. In the first part of the thesis, a network architecture establishing the feasibility of multi-user mmWave mobile communications is proposed. Stochastic geometric methods are used for characterising the performance of the system with considerations of the peculiar characteristics associated with high-frequency communication. In the second part, low-complexity algorithms are designed that result in energy-efficient mmWave systems critical for supporting an exploding number of device connections. However, real-life practice constraints, such as inaccurate knowledge of the channel over which signals are transmitted, and limitations of the hardware used to create a mmWave system, must be accounted for. Therefore, the third part investigates a new transmission method that compensates for some of the losses incurred by inaccurate channel knowledge while the fourth part examines the impact of specific hardware impairments on the performance of the mmWave systems, providing valuable design insights.

Contents

List of Figures

List of Tables

List of Abbreviations

1G	First Generation
3D	Three-Dimensional
4G	Fourth Generation
5G	Fifth Generation
ADC	Analogue-to-Digital Converter
AM	Amplitude Modulation
AoA	Angle of Arrival
AoD	Angle of Departure
ATN	Amplified Thermal Noise
AWGN	Additive White Gaussian Noise
BB	Baseband
BBU	Baseband Unit
BCP	Best Channel Participation
BS	Base Station
CAP	Continuous Aperture Phase
CCDF	Conditional Cumulative Distribution Function
CDF	Cumulative Distribution Function
CPRI	Common Public Radio Interface
CRAN	Cloud Radio Access Network
CS	Compressed Sensing
CSI	Channel State Information
CSIT	Channel State Information at the Transmitter
D2D	Device-to-Device
DAC	Digital-to-Analogue Converter
DFT	Discrete Fourier Transform
dB	Decibel
DoF	Degree of Freedom
EB	Exabytes

EE	Energy Efficiency
EIRP	Effective Isotropic Radiated Power
EM	Expectation Maximisation
EVM	Error Vector Magnitude
FCC	Federal Communications Commission
FEC	Forward Error Correction
FM	Frequency Modulation
GAMP	Generalised Approximate Message Passing
GHz	Gigahertz
I/Q	In-phase/Quadrature
IC	Integrated Circuit
ICI	Inter-Cluster Interference
IoT	Internet of Things
ITU	International Telecommunications Union
LNA	Low Noise Amplifier
LO	Local Oscillator
LOS	Line-of-Sight
LTE	Long-Term Evolution
MAC	Media Access Control
MGF	Moment Generating Function
MIMO	Multiple-Input Multiple-Output
MISO	Multiple-Input Single-Output
MMSE	Minimum Mean Squared Error
mmWave	Millimeter Wave
MUI	Multi-User Interference
NLOS	Non-Line-of-Sight
NNP	Nearest Neighbor Participation
NoRS	No Rate Splitting
Ofcom	Office of Communications
OFDM	Orthogonal Frequency Division Multiplexing
OMP	Orthogonal Matching Pursuit
PDF	Probability Distribution Function
PGFL	Probability-Generating Functional
PN	Phase Noise
PPP	Poisson Point Process
QoS	Quality of Service

RF	Radio Frequency
RRH	Remote Radio Head
RS	Rate Splitting
RATHI	Residual Additive Transceiver Hardware Impairment
RTHI	Residual Transceiver Hardware Impairment
RZF	Regularised Zero-Forcing
SDN	Software-Defined Networking
SE	Spectral Efficiency
SIC	Successive Interference Cancellation
SINR	Signal-to-Interference-Noise Ratio
SISO	Single-Input Single-Output
SNR	Signal-to-Noise Ratio
SOCP	Second-Order Cone Programming
SVD	Singular Value Decomposition
TV	Television
UE	User Equipment
ULA	Uniform Linear Array
UPA	Uniform Planar Array
V2I	Vehicle-to-Infrastructure
V2V	Vehicle-to-Vehicle
VCO	Voltage-Controlled-Oscillator
WLAN	Wireless Local Area Network
WMMSE	Weighted Minimum Mean Squared Error
w.p.	With Probability
WPAN	Wireless Personal Area Network
WSR	Weighted Sum Rate
ZF	Zero-Forcing

List of Notations

$\lvert \cdot \rvert$	Absolute value
$\lVert \cdot \rVert_0$	Zero norm
$\lVert \cdot \rVert_2$	Euclidean norm
$\lVert \cdot \rVert_F$	Frobenius norm
$(\cdot)^{-1}$	Inverse
$(\cdot)^H$	Hermitian transpose
$\otimes$	Kronecker product
a	Scalar
$\mathbf{a}$	Vector
$\mathbf{A}$	Matrix
$[\mathbf{A}]_m$	The mth element along diagonal of $\mathbf{A}$
$[\mathbf{A}\lvert\mathbf{B}]$	Horizontal concatenation
$\mathbb{C}$	Set of complex numbers
$\mathcal{CN}(a, b)$	Complex Gaussian random variable with mean a and variance b
$\mathrm{cov}(\cdot)$	Covariance
$\det$	Determinant
$\mathrm{diag}\{\mathbf{A}\}$	Diagonal matrix containing the elements along the diagonal of $\mathbf{A}$
$\mathbb{E}[\cdot]$	Expectation operator
$\mathrm{eigvec}(\mathbf{A})$	Eigenvectors of $\mathbf{A}$
$\exp(a)$	Returns e^a, where e is Euler's number
$\mathbf{I}$	Identity matrix
$\lim$	Limit
$\ln$	Base e logarithm
$\log_2$	Base 2 logarithm
$\max$	Maximise
$\min$	Minimise
$\mathcal{N}(a, b)$	Real Gaussian random variable with mean a and variance b
$\mathrm{tr}(\cdot)$	Trace

$\text{sgn}(\cdot)$	Signum function
$\text{vec}(\mathbf{A})$	Vector obtained by stacking columns of $\mathbf{A}$

Chapter 1

Introduction

This chapter begins by highlighting the real-world demands for improved wireless communication technology to provide unprecedented amounts of bandwidth. The unique contributions of this thesis are then provided, followed by an overview of the chapters included.

1.1 Background

The evolution of mobile communications over the past decade has played a significant role in transforming our modern societies. From the first-generation (1G) networks of voice-centric, analogue mobile phone systems to the latest data-centric fourth-generation long-term evolution networks (4G-LTE) deployed widely across the globe, mobile communications have radically changed the way information in contemporary society is accessed, shared and exchanged. Currently, with the rapid proliferation of social media, video streaming, mobile applications, and internet usage on smart devices (e.g. phones, tablets, watches, etc.), we are in an era of unprecedented demand for mobile data traffic, as illustrated by a 71% increase in global mobile data traffic between 2016 and 2017 [1]. Additionally, as depicted in Fig. 1.1, mobile data traffic is expected to grow at a compound annual growth rate of 46% between 2017 and 2022, with projections estimating an astonishing sevenfold increase in mobile data traffic of 77 exabytes (EB) per month in 2022 from 12 EB per month in 2017.

The rapid evolution of smart wireless devices has driven this exponential global increase in mobile connections from 8.6 billion connections in 2017 to an estimated 12.3 billion by 2022 [1]. This combination of increasing devices and the

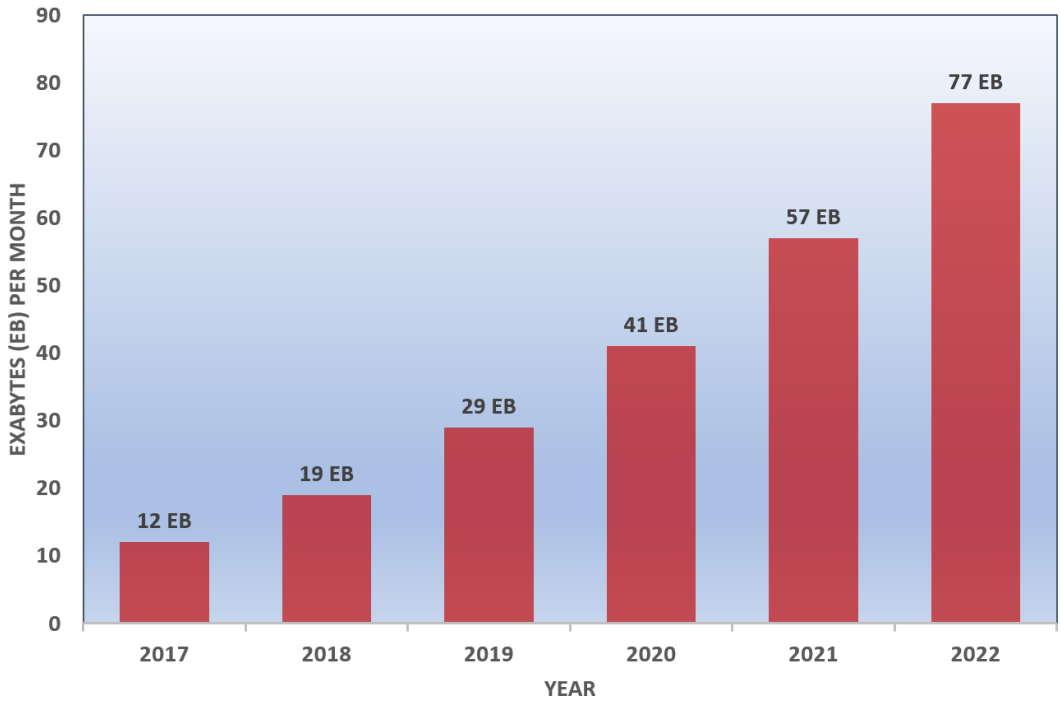


Figure 1.1: A projection of the global mobile data traffic, 2017–2022 [1].

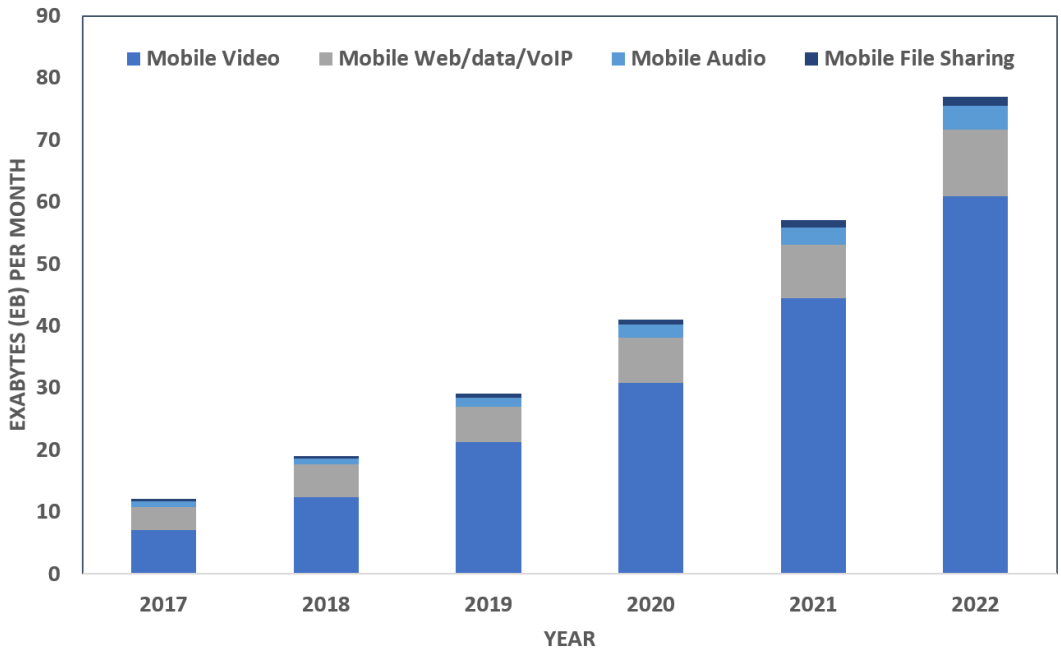


Figure 1.2: Content type distribution of global mobile data traffic [1].

increasing average data volume requested per device are the major contributors to the growth in mobile data traffic [2]. Fig. 1.2 shows the distribution of mobile data traffic in EB per month over four main mobile-content categories. It can be considered that about 79% of the world's mobile data traffic will be video by 2022. This calls for next-generation mobile communication systems to tackle the challenges brought by the consumer-led exponential growth in wireless data traffic.

Consequently, the development of fifth-generation (5G) networks is receiving significant interest from academia, industry and governments. The International Telecommunications Union (ITU) has stipulated 2020 as the year for standardising 5G mobile networks with the primary objectives of supporting billions of connected devices, providing data rates in gigabits per second for individual users, lowering end-to-end latencies for all connected devices and improving capacities and reliabilities relative to the existing 4G-LTE networks [3]. To this end, this thesis aims to contribute to the current research efforts of achieving these 5G goals with a focus on millimeter wave (mmWave) communications.

1.2 Contributions

The radio frequency (RF) spectrum, i.e., the part of the electromagnetic spectrum that can be used for wireless communication, has been globally controlled by the ITU[1] and divided into frequency bands dedicated to different radio technologies and applications such as satellite communication systems, TV broadcasting, maritime communications and military radar. Cellular communication systems today use specific frequencies on the RF spectrum typically under 6 GHz [4]. This part of the spectrum has become congested due to the ever-increasing growth in wireless devices. Thus, utilising an alternative spectrum for cellular communication has become an essential step to meet the objectives of 5G communication networks. In this thesis, we consider the mmWave spectrum range between 30–300 GHz, with an unparalleled amount of available spectrum for cellular communication. The specific contributions of the thesis can be summarized as follows

- We propose and study the performance of a multi-user system that cou-

[1]Different regions have governing bodies that control the allocation of frequency bands in their region; for example, the Office of Communications (Ofcom) in the UK and the Federal Communications Commission (FCC) in the USA. Notwithstanding, they allocate frequencies according to the ITU radio regulations.

ples the cloud radio access network (CRAN) architecture with mmWave. By modelling the remote radio heads (RRHs) and mobile user equipment (UEs) as independent poisson point processes (PPPs), we characterise the outage probability, average latency and throughput of this system under key factors, such as blockages, RRH density and path loss. The analysis is carried out for different user association configurations, namely, best channel participation (BCP) and nearest neighbour participation (NNP). Additionally, for both user configurations, we derive novel outage probability outer bounds in the interference-limited case, and closed-form expressions in the noise-limited case. Results show that (i) with high RRH deployment, larger antenna arrays can compensate for communication degradation (in terms of outage performance and latency), indicating a trade-off between inter-cluster interference (ICI) and RRH density, (ii) blockages and path loss tend to have a positive effect of decreasing outage probability in highly dense urban deployments and (iii) for user association configurations, BCP is the most viable for multi-user mmWave CRAN systems.

- We design hybrid precoders and combiners for the energy-efficient maximisation in a multi-user mmWave multiple-input, multiple-output (MIMO) scenario involving internet of things (IoT) devices. Since the original hybrid transceiver design problem is intractable, we propose a low-complexity two-stage algorithm, both stages of which are proven to converge. In the first stage, we transform the multivariate transceiver problem into its univariate equivalent, and then we exploit the well-known relationship between the weighted sum rate (WSR) and the weighted minimum mean squared error (WMMSE) to solve the univariate problem. In the second stage, we separate the joint transceiver optimization problems into hybrid precoder and combiner sub-problems and, using results obtained from stage one, we apply an orthogonal matching pursuit algorithm to derive the optimal solution for each sub-problem. We also take into consideration the transceiver hardware distortions that occur in practical systems and expand the analysis to cater for imperfect channel state information (CSI). Results show that the proposed hybrid algorithms achieve near-optimal energy efficiency (EE) performance to support power-constrained IoT devices with lower complexity than the univariate designs, and are resilient to errors obtained from imperfect channel estimation.

- We consider the use of a rate-splitting (RS) transmission scheme to mitigate the detrimental effects of interference in a multi-user mmWave network with hybrid beamforming under imperfect CSI. Within this context, the transmitted signal for each user is divided into a common message and a private message. We dynamically split the power between the common and private messages such that there is interference-free transmission for the common message. Additionally, we analyse the sum rate of RS and optimize the precoders of the common messages. Simulation results show that RS can tackle the residual multi-user interference (MUI) caused by CSI mismatch in the mmWave system, achieving a sum-rate gain over linear transmission strategies. Moreover, RS is advantageous for the practical implementation of mmWave systems with constrained hardware since it mitigates the detrimental effects of the transceiver impairments on the performance of the system. Furthermore, greater efficiency of RS can be obtained by increasing the resolution parameters for channel estimation.

- We consider the spectral efficiency (SE) performance of a mmWave MIMO system that generalises the state of the art by considering the inevitable residual transceiver hardware impairments (RTHIs). We shed light on the impact of three major hardware impairments, namely, the multiplicative phase noise (PN), the amplified thermal noise (ATN) and the residual additive transceiver hardware impairments (RATHIs). In particular, we design the hybrid beamformers that maximise the SE of the system taking into consideration the effects of the RTHIs. Additionally, we quantify the degradation caused to the spectral efficiency of the system by each separate impairment. Results show that the multiplicative PN is the most critical of the three impairments. Furthermore, we verify the merits of using massive MIMO in mmWave communications to provide gains that reduce the dependence of the received signal power on the signal wavelength whilst allowing for a cost-efficient deployment of the system.

1.3 Thesis Outline

The rest of this thesis is organised as follows.

- Chapter 2 provides an overview of relevant concepts in mmWave wireless communications to prepare the reader for the technical subjects covered in

the thesis, which include mmWave channel and blockage modelling, different MIMO architectures applicable for mmWave transmission and channel estimation strategies for the practical implementation of mmWave systems.

- In Chapter 3, we design a multi-user mmWave cloud CRAN system and analyse its performance using the metrics of outage probability, average latency and throughput. We consider two transmission schemes based on how the user is associated with the transmitter and derive closed-form expressions of the outage probability under the noise-limited scenario. Additionally, bounds of the outage probability are derived for the interference-limited scenario.

- Chapter 4 focuses on the design of EE hybrid beamformers in a multi-user mmWave system involving IoT devices under perfect and imperfect CSI scenarios. The EE maximisation problem is solved by using the relationship between WSR and WMMSE and an alternating maximisation algorithm after the separation of the joint transceiver problem into separate hybrid precoder and combiner sub-problems. Numerical results validate the hybrid beamforming design and demonstrate its effectiveness in addressing the energy-efficiency concerns in the implementation of practical mmWave systems.

- Chapter 5 studies the performance of a multi-user mmWave system under the imperfect CSI. We analyse the sum rate by proposing an RS transmission strategy, which separates the message for each user in the system into a common part and a private part. We also take into consideration the limited dynamic range of practical transmitters and receivers and use optimization techniques to demonstrate that RS achieves significant gains over the conventional linear transmission strategies.

- In light of the practical limitations in assuming ideal hardware for mmWave MIMO systems, Chapter 6 addresses the problem of assessing the SE performance of the mmWave system under three major residual hardware impairments: (a) multiplicative PN, (b) RATHIs and (c) ATN. We develop a framework for the modelling of RTHIs and within this context measure their impact on the mmWave system.

- Chapter 7 concludes the thesis by summarising the main findings of this research, discussing its limitations and providing suggestions for future work.

Chapter 2

Background

This chapter provides background material on the fundamentals of mmWave wireless communication that supports the technical chapters presented in the thesis. First, we present the mmWave spectrum, outlining its distinctive features, advantages and applications. Next, we present a brief overview of mmWave propagation, characterising its channel and blockages with statistical models. This is relevant to Chapters 3, 4, 5 and 6. Then, we focus on MIMO architectures at mmWave, introducing a variety of beamforming strategies; this will play an important role in Chapters 3–6. Finally, we consider channel estimation strategies which will support the work presented in Chapters 4 and 5.

2.1 Millimeter Wave Spectrum

Mobile communication has become an essential part of the modern world. With the explosive proliferation of mobile wireless services in every sector of society, and the rapid worldwide development of wireless devices in various forms that require high data rates, mobile data traffic and connections will show unprecedented growth in the future.

The radio spectrum is one of the most important resources for mobile communications. Until recently, the spectrum assigned for wireless systems has been under 6 GHz,[1] which has been centrally controlled (by Ofcom in the UK and the FCC in USA) using a policy that divided the assigned spectrum into frequency bands dedicated to a particular service and guaranteeing a specific level

[1]Throughout this thesis, wireless systems operating at frequencies under 6 GHz will be referred to as "standard wireless systems".

of service to the licensed users. More recently, an adaptive approach of spectrum usage was adopted, which allows cognitive radios to dynamically sense the spectrum and make use of idle frequency bands [5]. Despite this efficient use of the existing spectrum, the available bandwidth is not commensurate with the increase in demand. To this end, there has been increased interest in standardising the spectrum above 6 GHz for mobile communications.

The mmWave spectrum (30–300 GHz), currently used for radar, military and backhaul communication, has a much lower utilisation level. Fig. 2.1 illustrates the available spectrum for use between 30–100 GHz. Notably, the unlicensed band at 60 GHz surpasses all other bands in the spectrum that have been used for cellular, satellite, WiFi, AM/FM radio and TV communication in the world! [4]. The staggering amount of bandwidth available makes mmWave a frontier for wireless communication systems.

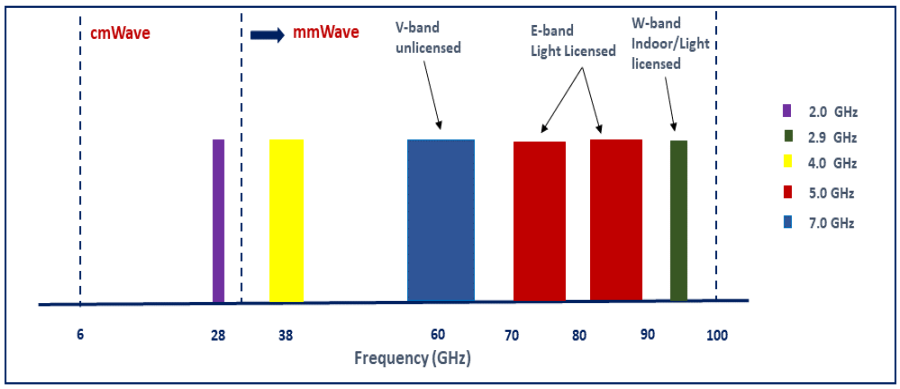


Figure 2.1: An illustration of the potential bandwidth available in the mmWave spectrum [71].

2.1.1 Distinctive mmWave characteristics

The distinct characteristics of high-frequency propagation imply that the key technologies used at lower frequencies cannot directly be applied in mmWave communications. The main distinctive features which are imperative for modelling mmWave systems and understanding signal propagation at these frequencies are discussed as follows.

- *Isotropic Path Loss*: The transmitted signal in all wireless communication systems undergoes distance-based attenuation. According to Friis Law, the isotropic path loss increases in direct proportion to the inverse of the squared wavelength i.e, $\frac{P_{\mathrm{TX}}}{P_{\mathrm{RX}}} \propto \lambda^{-2}$ [6]. This implies that mmWave propagation will experience higher path loss when compared to standard systems transmitting at lower frequencies. However, directional transmissions can compensate for this high path loss through directional antenna gains [7]. Thus, large arrays are required to provide these gains. Fortunately, realising large antenna arrays at both the transmitter and receiver of mmWave systems is possible due to the small wavelengths at these high frequencies. Accordingly, MIMO antenna arrays are integral to providing quality mmWave communications.

- *Received signal power*: In a standard wireless system, the aperture of the receive antenna is dependent on the carrier frequency and the received signal power is given by

$$P_{\mathrm{RX}} = \frac{P_{\mathrm{TX}}}{4\pi R^2}\frac{\lambda^2}{4\pi}, \tag{2.1}$$

where P_{RX} and P_{TX} are the receive and transmit power, R is the separation between the transmitter and receiver and λ is the wavelength. However, given that high frequencies translate to small wavelengths, this would result in very small received power for a mmWave system, as illustrated in Fig. 2.2. Hence, to achieve quality communication, there is a necessity to reduce the dependence of the received signal power on the wavelengths of the carrier frequencies in mmWave systems. Fortunately, this can be achieved by keeping the aperture constant with the use of large antenna arrays for transmit and receive beamforming as illustrated in Fig. 2.3. Within this context, the received power is expressed as

$$P_{\mathrm{RX}} = \underbrace{\frac{P_{\mathrm{TX}}}{4\pi R^2}}_{\substack{\text{Receive}\\ \text{Spectral}\\ \text{Density}}} \underbrace{\frac{\lambda^2 G_{\mathrm{RX}}}{4\pi}}_{\substack{\text{Effective}\\ \text{Receive}\\ \text{Aperture}}} \underbrace{G_{\mathrm{TX}}}_{\substack{\text{Transmit}\\ \text{Antenna}\\ \text{Gain}}}. \tag{2.2}$$

Accordingly, equation (2.2) further motivates the use of large antenna arrays in mmWave communication. For example, the array sizes of 64 elements [8] and 256 elements [9] have been used in experiments to show the feasibility of mmWave communications.

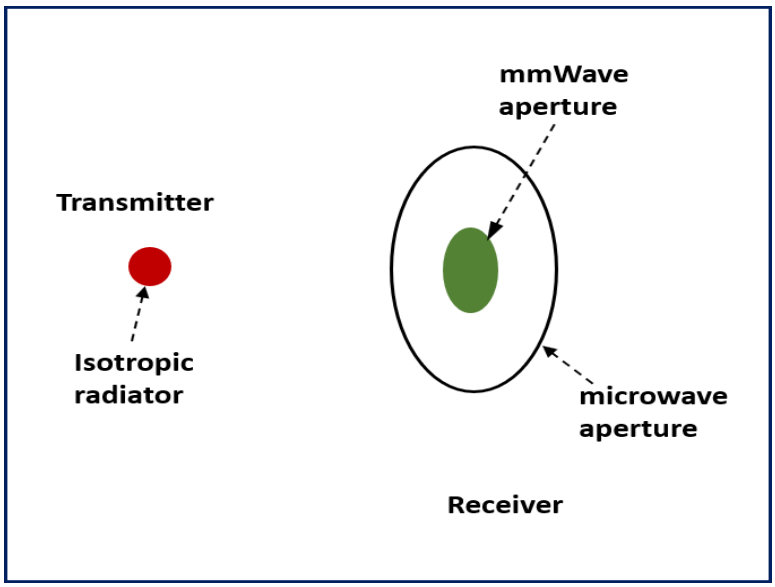


Figure 2.2: An illustration of a typical receiver dependent on carrier frequency.

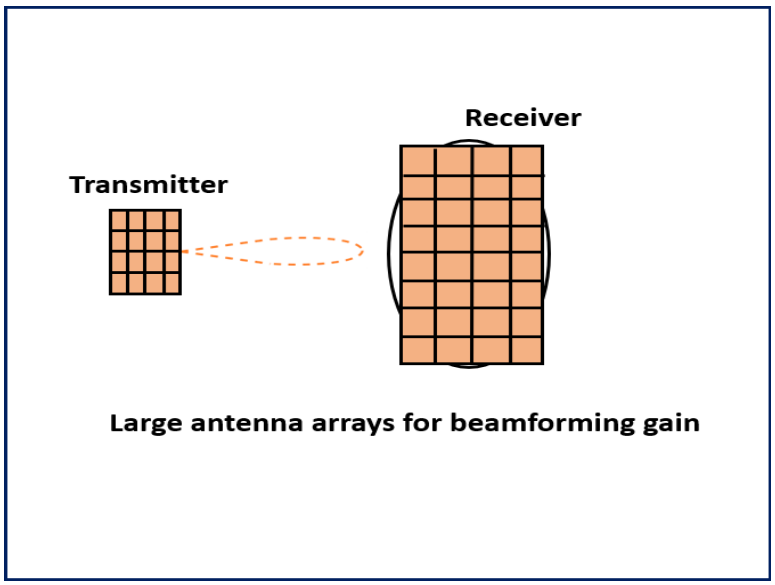


Figure 2.3: An illustration of a basic mmWave system with large antenna array.

- *Signal absorption and blockages*: Compared to standard wireless systems, mmWave signals are more susceptible to blockages since most objects in a propagation environment(e.g., walls, lampposts, birds, people, etc.) are larger than the signal wavelength. Nevertheless, reflection and scattering allow for transmission between a transmitter and a receiver when steerable antennas are used to locate objects that reflect or scatter energy [4]. Thus, antenna arrays capable of directional and adaptive beamforming become

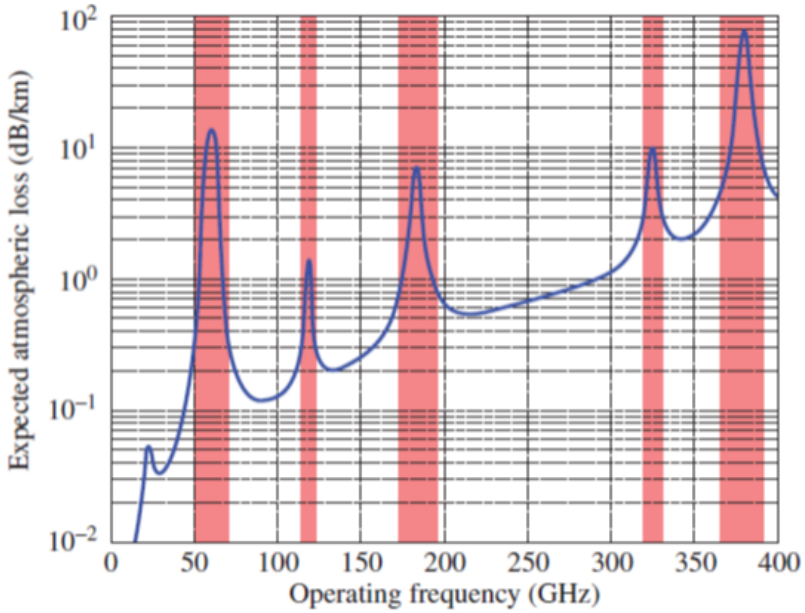


Figure 2.4: An illustration of the excess attenuation achievable across 0–400 GHz frequency bands [4].

essential for successful communication in mmWave systems [10]. Furthermore, mmWave signals are affected by atmospheric conditions such as temperature and humidity, as they impact the molecular components of air and water which can cause atmospheric absorption of small wavelength signals. Fig. 2.4 illustrates the attenuation in addition to free space loss across the mmWave spectrum. It can be observed that there is severe attenuation at certain spectrum bands i.e., 60 and 180 GHz. Consequently, these frequencies will be suitable for short-range wireless technologies. Moreover, there is little attenuation (below 1 dB/km) in the 30–50, 70–90, 120–160 and 200–300 GHz bands, making them well-fitted to longer distances and suitable candidates for future mobile and cellular communications. Additionally, adaptive antenna arrays can be used to compensate for specific atmospheric losses as they can be adapted to adjust to particular levels of interference in the low-attenuation bands [4].

- *Sparse channels*: When compared to standard wireless systems, mmWave channels tend to be sparse in the number of significant scatterers that contribute to the multi-path components of a transmitted signal [11]. This has been verified by measurements of the 28- and 73-GHz channels in [12]. The results show that the mmWave channel have 2 clusters on average con-

tributing a propagation path, with the location of the angles of arrivals close to the boresight direction. The sparsity in mmWave channels asserts the necessity of alternate channel models that are different from existing standard wireless channel models.

2.1.2 Advantages of mmWave Communication

The advantages of mmWave can be quantified by the solutions it proffers for future 5G cellular networks. Compared to the existing 4G-LTE networks, mmWave offers a number of remarkable advantages, as follows [13].

- *Bandwidth allocations*: Evidently, mmWave carrier frequencies (in the order of GHz) allow for larger channel bandwidth allocations which directly result in higher data rates and indirectly to reduced latencies as insufficient bandwidths reduce the latency of the network. Accordingly, service providers will be able to support user-data-hungry applications with minimal latency.

- *Capacity gain*: In addition to expanding the channel bandwidth, mmWaves can be used to reduce coverage areas, i.e., establish more densely packed communication links and exploit spatial reuse to provide increased capacity gains.

- *Homogeneity*: Unlike the disjointed spectrum (i.e., a high variance between the frequencies in MHz and GHz (over three octaves of frequency)) utilized by operators of standard wireless systems, the spectral allocation in the mmWave spectrum is likely to be relatively close together, implying comparable propagation characteristics between different mmWave bands [13].

Additionally, the performance gains of utilising MIMO over single-input single-output (SISO) in standard wireless systems are also applicable to mmWave communications since mmWave frequencies may exploit massive MIMO techniques. These gains include the following.

1. *Interference suppression*: In multi-user systems, the use of multiple antennas at the transmitter and receiver increase the potential to alleviate

intra-channel interference. This is achieved via precoding at the transmitter or combining at the receiver, or a combination of both. However, the precoding technique utilised in mmWave systems differs from that employed in standard transmissions; this will be discussed in detail in section 2.3.

2. *Diversity gain*: Spatial diversity can be exploited in mmWave systems with multiple antennas at both ends to mitigate the impact of fluctuations in the channel and loss of signals.

3. *Multiplexing gain*: With multiple antennas at the transmitter, parallel streams can be transmitted to the user without using additional bandwidth or power. This increases the number of spatial dimensions for communication.

2.1.3 Challenges of mmWave Communication

Despite the great potential associated with mmWave communications as outlined in section 2.1.2, a number of challenges need to be addressed to exploit these benefits. These will be discussed as follows

- *Power consumption*: mmWave systems require large antenna arrays to reduce the dependence of received power on the small signal wavelengths and for implementing directional beamforming. A direct consequence of this is an increase in processing power consumption. This is a significant challenge as the power consumed by devices essential for signal processing such as data converters, mixers, and power amplifiers scale up with the increasing number of antennas. For example, 12-bit ADCs in a mmWave system with 16 antennas would consume more than 250mW [14]. This implies that alternate signal processing techniques are required for mmWave systems and motivates the hybrid architecture explained in section 2.3.2 and implemented in all the technical chapters of the thesis.

- *Integrated circuit (IC) design*: Additional factors that need to be considered when designing ICs for mmWave systems with high carrier frequencies and wide bandwidth include non-linear distortions in the power amplifiers, phase noise and IQ imbalance because the severity of these errors scale up with high frequency transmissions [15].

- *User Mobility*: A major challenge that comes with user mobility in mmWave transmissions is the significant fluctuations of the channel state since channel coherence time in the mmWave range is very small resulting in a large doppler spread [14]. Thus, the modulation and coding schemes used in mobile mmWave communication must take into account the varying channel states and motivates the need for channel estimation in mmWave cellular networks. Additionally, in indoor environments, user mobility cause rapid load fluctuations requiring optimized load balancing schemes between access points [16].

2.1.4 Applications of mmWave Communication

As mentioned earlier in section 2.1, the mmWave spectrum is already in use for military, radar and wireless backhaul applications. The first standardised commercial applications of mmWave technology were provided by wireless personal area networks (WPAN) and wireless local area networks (WLAN) using the short-ranged 60-GHz unlicensed bands. WirelessHD, being the most popular standard of WPAN, provided a high-bandwidth interface as a replacement for cables carrying uncompressed high-definition multi-media content [17]. IEEE 802.11ad is the standard for WLAN incorporated in laptops, tablets and smartphones to connect these devices to the internet via a wireless access point [7]. Notwithstanding, mmWave technology is promising for many potential applications, discussed as follows.

- *5G communication systems*: Although mmWave has been established for providing gigabit data in the wireless backhaul links of communication systems, the traditional design required expensive components to provide reliable long-ranged communication links, reducing the cost advantages over wired backhaul solutions. However, with the increase in density of cellular systems, especially in urban environments, low-cost mmWave technologies are actively being researched and developed to offer high-capacity backhaul links for densely distributed cellular networks [18, 19]. Furthermore, the development of mmWave is not limited to low-cost, backhaul links; it also has great potential to handle fronthaul and access links. This application for 5G cellular systems is receiving tremendous research interest [11, 12, 13, 20, 21, 22, 23] because of the potential gains (outlined earlier in section 2.1.2) possible with mmWave transmissions. To this end, Chapters

3, 5 and 6 of this thesis will address the performance of mmWave communications for cellular networks.

- *Vehicular networks*: Recently, there has been a growing interest in vehicular applications to unlock many social benefits. For example, autonomous vehicles, vehicle-to-vehicle (V2V) communication, and vehicle-to-infrastructure (V2I) links may be used for the exchange of traffic information, avoidance of collision and pollution management[24]. Given that mmWave is already established in automotive radar applications [25], one can naturally expect mmWave to play a significant role in supporting vehicular networks.

- *Wearable networks*: Technology advancement is enabling the rapid evolution of wireless devices. This is driving an increasing interest in wearable networks which will provide connections for smart devices such as cellphones, tablets, watches, activity-tracking devices, virtual-reality headsets and glasses. These high-end devices require high data rate and very low latencies to communicate with one another [26]. Consequently, mmWave is of interest to these networks for its potential to meet their requirements.

- *IoT networks*: IoT is an emerging paradigm that provides connectivity of physical things in a network, impacting all sectors of the society. Take the educational sector, for example; IoT will enable transfer of information (e.g entire libraries) from lampposts to students in building hallways or entrances, thus eliminating the need to carry physical textbooks between classes [27]. This will create a huge network comprising billions of connections, and the massive mmWave spectrum has the capacity to support such connections at high data rates. This makes it foreseeable that mmWave will be fundamental in enabling IoT for the future. Accordingly, Chapter 4 of this thesis will investigate mmWave communication involving IoT devices.

2.2 Millimeter Wave Propagation

The basic form of mmWave communications includes a point-to-point link where a single transmitter equipped with a large array of antennas communicates with one multi-antenna receiver. The transmitted signal is affected by its surrounding environment (highlighted earlier in section 2.1.1), and experiences multiple paths due to scatterers and reflectors, resulting in multiple distorted versions of

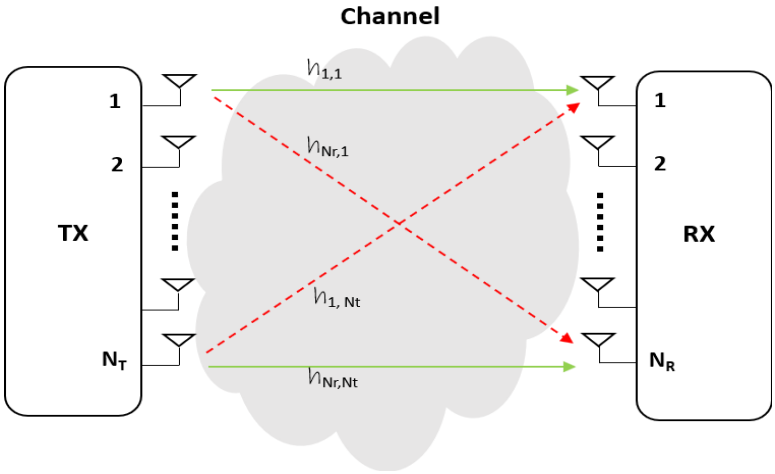


Figure 2.5: An illustration of a basic $N_{\mathrm{R}} \times N_{\mathrm{T}}$ mmWave system.

the transmitted signal arriving at the receiver, (These propagation effects are commonly referred to as "small-scale fading"). In addition to the small-scale channel effects, each multi-path version of the signal experiences distance-based attenuation and a shadowing loss from obstacles (known as large scale fading). These propagation effects as well as blockages form the fundamental aspects of mmWave communications. Fortunately, MIMO antenna arrays are integrated into mmWave communication, and these large arrays can be exploited to achieve directional transmissions through directional beamforming which both compensates for the distance-based path loss and can manage interference through advanced techniques of beam shaping [28]. Fig. 2.5 illustrates a mmWave system with N_{T} transmit antennas and N_{R} receive antennas. The received signal of the system is given by

$$\mathbf{y} = \mathbf{H}\mathbf{s} + \mathbf{z}, \tag{2.3}$$

where $\mathbf{y} = [y_1, y_2, \ldots, y_{N_{\mathrm{R}}}]^T$ is the signal vector received by N_{R} antennas, $\mathbf{s} = [s_1, s_2, \ldots, s_{N_{\mathrm{T}}}]^T$ is the signal vector transmitted by N_{T} antennas, $\mathbf{H}$ is the $N_{\mathrm{R}} \times N_{\mathrm{T}}$ mmWave channel matrix and $\mathbf{z} = [z_1, z_2, \ldots, z_{N_{\mathrm{R}}}]^T$ is the additive white Gaussian noise vector.

2.2.1 Channel models

2.2.1.1 Narrowband channel model

In the case of mmWave propagation, the channel presents a number of characteristics (outlined in section 2.1.1) that differentiate it from the standard MIMO

channel. Hence, the statistical fading distributions employed in latter cannot be directly applied for mmWave communications. With these characteristics, the Saleh–Valenzuela model extended to mmWave channels [29, 30] enables us to describe the structure of a mmWave channel mathematically by means of a narrowband clustered channel model. From a physical point of view, it represents a geometric model with N_{p} scatterers, where each scatterer is assumed to contribute a signal propagation path given as

$$\mathbf{H} = \gamma \sum_{l=1}^{N_{\mathrm{p}}} \alpha_l \, \mathbf{a}_{\mathrm{R}} \left(\phi_{\mathrm{R}}^{l}, \theta_{\mathrm{R}}^{l}\right) \mathbf{a}_{\mathrm{T}} \left(\phi_{\mathrm{T}}^{l}, \theta_{\mathrm{T}}^{l}\right)^{H}, \tag{2.4}$$

where $\gamma = \sqrt{\frac{N_{\mathrm{T}} N_{\mathrm{R}}}{N_{\mathrm{p}}}}$ denotes the normalization factor. Each path l is described by its complex gain α_l, the azimuth and elevation angles of departure $(\phi_{\mathrm{T}}^{l}, \theta_{\mathrm{T}}^{l})$, and the corresponding angles of arrival $(\phi_{\mathrm{R}}^{l}, \theta_{\mathrm{R}}^{l})$, $\mathbf{a}_{\mathrm{R}}(\phi_{\mathrm{R}}^{l}, \theta_{\mathrm{R}}^{l})$ and $\mathbf{a}_{\mathrm{T}}(\phi_{\mathrm{T}}^{l}, \theta_{\mathrm{T}}^{l})^{H}$, denote the steering vectors of the transmitter and receiver respectively.

For an N-element uniform planar array (UPA), the antenna array response vector is given as

$$\mathbf{a}(\phi, \theta) = \frac{1}{\sqrt{N}} \left[1, e^{j2\frac{\pi}{\lambda} d(a \sin(\phi) \sin(\theta) + b \cos(\theta))} \ldots, e^{j2\frac{\pi}{\lambda} d((A-1) \sin(\phi) \sin(\theta) + (B-1) \cos(\theta))}\right], \tag{2.5}$$

where $0 \leq a \leq A$ and $0 \leq b \leq B$ are the indices of antenna elements with $N = AB$, λ is the wavelength and d is the distance between the antenna elements. In 2D channels, the array geometry is linear. In such cases, the array steering vectors are described by uniform linear arrays (ULA). Hence, an N-element ULA is given as

$$\mathbf{a}(\phi) = \frac{1}{\sqrt{N}} \left[1, e^{j2\frac{\pi}{\lambda} d \sin(\phi)} \ldots, e^{j2\frac{\pi}{\lambda} d(N-1) \sin(\phi)}\right]. \tag{2.6}$$

Furthermore, the combination of the high dimensionality and the highly directional nature of mmWave propagation means that the narrowband channel can alternatively be represented by its beamspace [7]. In other words, the array steering vectors can be used to define the communication space of the channel. Accordingly, equation (2.4) can be rewritten as

$$\mathbf{H} = \mathbf{A}_{\mathrm{R}} \mathbf{H}_{\alpha} \mathbf{A}_{\mathrm{T}}^{H}, \tag{2.7}$$

where $\mathbf{A}_{\mathrm{R}} \in \mathbb{C}^{N_{\mathrm{R}} \times N_{\mathrm{p}}}$, and $\mathbf{A}_{\mathrm{T}} \in \mathbb{C}^{N_{\mathrm{T}} \times N_{\mathrm{p}}}$ are the array response vectors of the

receiver and transmitter, respectively, and $\mathbf{H}_\alpha = \text{diag}(\boldsymbol{\alpha})$ contains the complex gains of all paths with $\boldsymbol{\alpha} = \left[\alpha_1, \alpha_2, \ldots, \alpha_{N_\text{P}}\right]$. The channel in equation (2.7) is commonly referred to as the virtual channel representation [31] and is useful for channel estimation. Since the beamspace and antenna space are related by a spatial Fourier transform, the virtual channel can also be expressed as [32]

$$\mathbf{H} = \mathbf{U}_\text{R} \mathbf{H}_\alpha \mathbf{U}_\text{T}^H, \tag{2.8}$$

where $\mathbf{U}_\text{R} \in \mathbb{C}^{N_\text{R} \times N_\text{R}}$ and $\mathbf{U}_\text{T} \in \mathbb{C}^{N_\text{T} \times N_\text{T}}$ are unitary discrete Fourier transform (DFT) matrices. The beamspace of an N-dimensional ULA is represented by spatial angles, ϕ_i, $i = 0 \ldots, N-1$ with uniform spacing and the corresponding array response vectors results in an orthonormal basis for the signal space. Specifically, the unitary DFT matrix is expressed as [31]

$$\mathbf{U} = \frac{1}{\sqrt{N}} \left[\mathbf{a}(\phi_0), \ldots, \mathbf{a}(\phi_{N-1})\right]. \tag{2.9}$$

2.2.1.2 Wideband channel model

For wideband operations in non-line-of-sight environments, the impact of the multi-path delay spread is more significant. Hence, each path is characterised by a time delay in addition to the complex gain and angles of departure and arrival [33]. Accordingly, the wideband channel model is used to incorporate the multi-path delays and is given as [34]

$$\mathbf{H}_\delta = \sum_{l=1}^{N_\text{P}} \alpha_l \, p_{rc}\left(\delta T_s - \tau_l\right) \mathbf{a}_\text{R}\left(\phi_\text{R}^l, \theta_\text{R}^l\right) \mathbf{a}_\text{T}\left(\phi_\text{T}^l, \theta_\text{T}^l\right)^H, \tag{2.10}$$

where $\mathbf{H}_\delta$ denotes the channel response at delay δ, $p_{rc}(\tau)$ is the raised cosine pulse-shaping filter generating pulses of the signal at T_s intervals evaluated at τ seconds, and τ_l is the time delay of the lth path. Note that the channel characteristics are assumed to be constant within the coherence time of $t = \delta T_s$ seconds.

2.2.2 Blockage models

Blockage models are integral for an accurate depiction of mmWave propagation since most objects can either attenuate or scatter mmWave signals. In cellular systems, a three-state model is used to measure the effects of blockages in the

channel. The best state is the line-of-sight (LOS), which occurs when there is a direct link between the transmitter and receiver with no obstacles between them, while the non-line-of-sight (NLOS) state occurs when the materials of the objects are reflective, causing an indirect path between the source and destination. The worst case is the signal outage state, when the obstacle significantly attenuates the transmitted signal so that it does not arrive at the receiver. Recent mmWave propagation measurements show that LOS and NLOS channels demonstrate different small-scale and large-scale fading characteristics [12, 13]. Two methods of modelling the impact of blockages in mmWave networks are (a) using the theory of ray tracing based on geographical information of the size, location and orientation of the blockages in the mmWave multi-path channel [11] and (b) using statistical models which describe the blockages according to a distribution [35]. The statistical approach of modelling blockages is utilised in this thesis because it can accommodate varying blockage parameters including density and size. The most common statistical models to measure the impact of blockages such as buildings are summarized as follows.

- The random shape theory model, commonly referred to as the "exponential blockage model" [35]: this model is applicable in both indoor and outdoor scenarios and models blockages as a randomly distributed process, as illustrated in Fig. 2.6. Using the Boolean scheme where building centres form a PPP, the probability that a link of length r is LOS is given by

$$P_{\mathrm{L}}(R) = e^{-\beta r}, \tag{2.11}$$

  where β is the blockage parameter, given as

$$\beta = \frac{-P\ln(1-\lambda)}{\pi A}, \tag{2.12}$$

  with A, λ and P denoting the average building area, the percentage coverage area of the building and the average building perimeter, respectively.

- The LOS ball model (also called the "fixed blockage model"): this is a simplified probabilistic blockage model verified by measurements in [18] and [14]. Here, the LOS probability is modelled as a step function; in other words, the link is assumed to be LOS on the probability that its length falls within a fixed radius, otherwise, the link is NLOS, as depicted in Fig. 2.7. Specifically, for a link of length r and a fixed radius R, the LOS probability

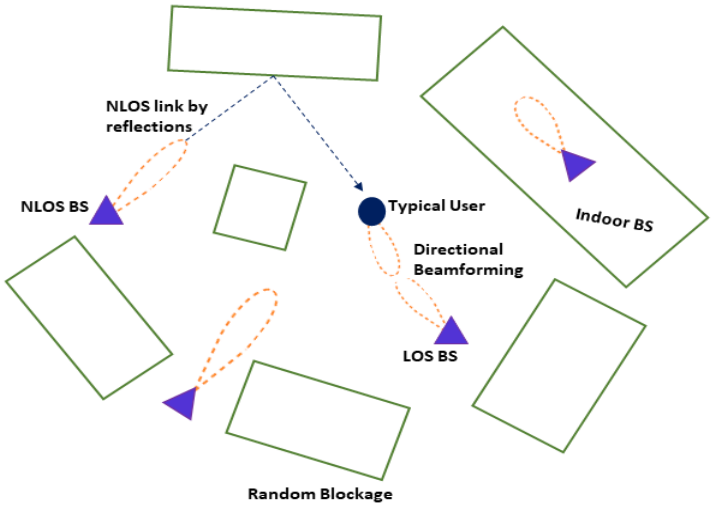


Figure 2.6: An illustration of building blockages modelled as a Boolean scheme of rectangles [10].

is given as [18]

$$P_{\mathrm{L}}(r) = \begin{cases} C, & \text{if } \ r < R \\ 0, & \text{otherwise,} \end{cases}, \tag{2.13}$$

where $0 \leq C \leq 1$, and $R > 0$.

- 3GPP model [36]: this model is dependent on the propagation environment and maps a specific distance to the probability of the link being LOS as 0 or 1. Specifically, for a typical outdoor urban setting, the LOS probability of a link of length r is given by

$$P_{\mathrm{L}}(r) = \min\left(\frac{18}{r}, 1\right)\left(1 - \exp\left(-\frac{r}{36}\right)\right) + \exp\left(-\frac{r}{36}\right), \tag{2.14}$$

whereas in a typical suburban area, it is given by

$$P_{\mathrm{L}}(r) = \exp\left(-\frac{d}{200}\right). \tag{2.15}$$

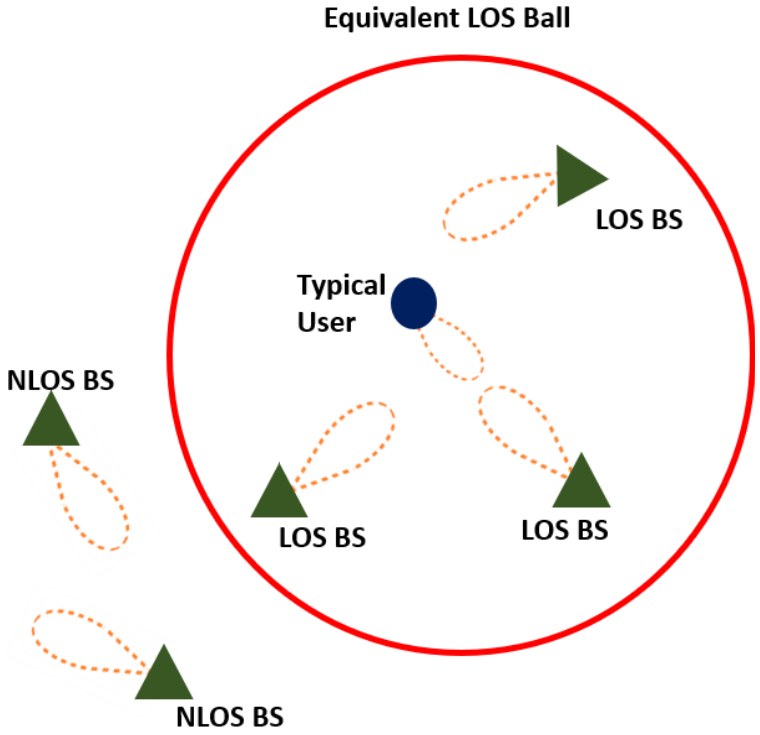


Figure 2.7: An illustration of building blockages modelled as a fixed ball model [10].

2.3 MIMO Architectures at mmWave

The MIMO architecture for standard wireless communications, where all the signal processing happens at baseband level, is illustrated in Fig. 2.8. From the figure, the received signal for a narrowband channel model is given by

$$\mathbf{y} = \mathbf{W}^H \mathbf{H} \mathbf{V} \mathbf{x} + \mathbf{W}^H \mathbf{z}, \tag{2.16}$$

where $\mathbf{V}$ denotes the $N_\mathrm{T} \times N_\mathrm{s}$ precoding matrix used to transmit the $N_\mathrm{s} \times 1$ symbol vector $\mathbf{x}$, and $\mathbf{W}$ represents the $N_\mathrm{R} \times N_\mathrm{s}$ combining matrix. A key characteristic in the existing commercial MIMO standards is that the technology required supports a small number of antennas (typically 2–8 [7]), as baseband processing requires a separate RF chain including data converters and signal mixers for every antenna. However, given that mmWave systems facilitate the use of tightly packed antenna arrays (typically 32–256 array elements [9]), the prohibitive costs in terms of complexity and power consumption make baseband processing infeasible for communications at mmWave frequencies. To quantify the impact of

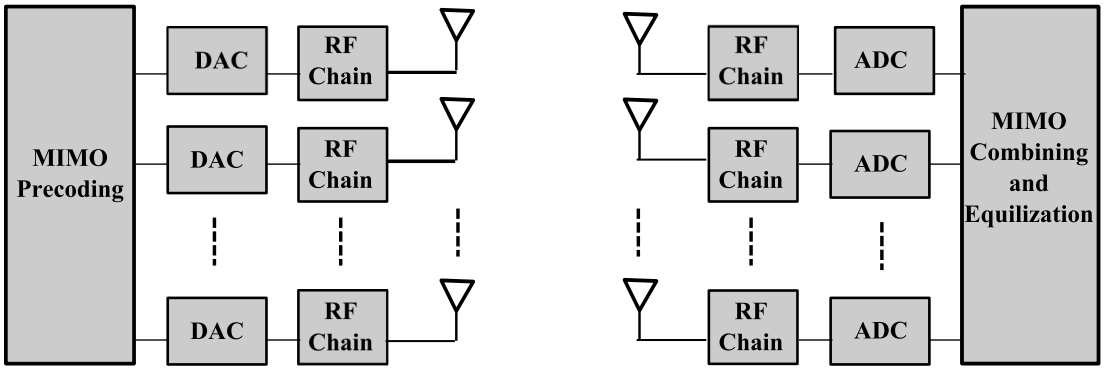


Figure 2.8: MIMO architecture in standard wireless systems.

power consumption in mmWave systems, Table 2.1 provides a comparison of the power consumption range of components in the receiver architecture between a mmWave system and a standard system where $N_{\mathrm{R}}^{\mathrm{RF}}$ represents the number of RF chains at the receiver, Comp. refers to Components, Req. to Required, Amp. to Amplifier, Conv. to Converter, LNA to Low-Noise-Amplifier, and VCO to Voltage-Controlled-Oscillator.[2]. Note that the exact power consumption depends on the technology and specification used to implement the device. Taking into account these hardware constraints, the following architectures have been proposed for application in mmWave systems.

2.3.1 Analog beamforming

The analog solution, proposed to replace baseband processing, is the simplest method for applying MIMO techniques in mmWave systems and requires only one RF chain for the entire antenna array. Analog beamforming is implemented in the RF domain using a network of phase shifters with quantised phases, as illustrated in Fig. 2.9. Here, different weights are dynamically assigned to the phase shifters which can steer the transmitted beam in different directions. Thus, efficient beam steering strategies are required to achieve maximum received power for the user. For example, in [19], the RF beamforming vectors are designed using an efficient beam alignment technique which uses adaptive subspace sampling and multi-resolution beam codebooks while in [31] the concept of beamspace MIMO was exploited to direct the transmitted signals to subspaces that maximise the

[2]The data of Table 2.1 was taken from recent studies showcasing prototype devices for mmWave systems[4, 7] and standard sytems[37, 38, 39, 40].

Component	mmWave System		Standard System	
	Comp. Req.	Power (mW)	Comp. Req.	Power (mW)
Phase Shifters	$N_{\mathrm{R}} \times N_{\mathrm{R}}^{\mathrm{RF}}$	15–110	N/A	N/A
Power Amp.	N_{R}	40–250	N_{R}	20–246
Data Conv.	$N_{\mathrm{R}}^{\mathrm{RF}}$	15–795	N_{R}	40–389
LNA	N_{R}	4–86	N_{R}	5.9–20
VCO	$N_{\mathrm{R}}^{\mathrm{RF}}$	4–25	N_{R}	0.15–67.5

Table 2.1: Power consumption range for different components of a receiver architecture.

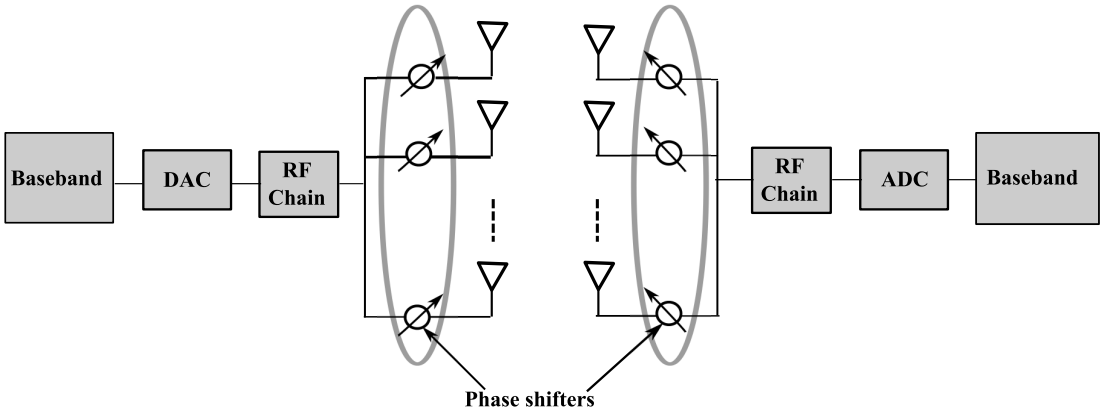


Figure 2.9: Analog beamforming using a network of phase shifters.

received signal power. Furthermore, analog beamforming has been adopted as a solution for commercial indoor mmWave communication standards such as IEEE 802.11ad [41].

Nevertheless, analog beamforming in mmWave systems is subject to several constraints. For example, the power consumed by the phase shifters is dependent on the phase angle resolution, and the performance is limited by non-linearities in the phase shifters and use of quantised phase shifts. These constraints limit the application of analog beamforming to single-stream and single-user transmissions, reducing the multi-stream and multi-user advantages associated with MIMO transmissions [7, 42].

2.3.2 Hybrid beamforming

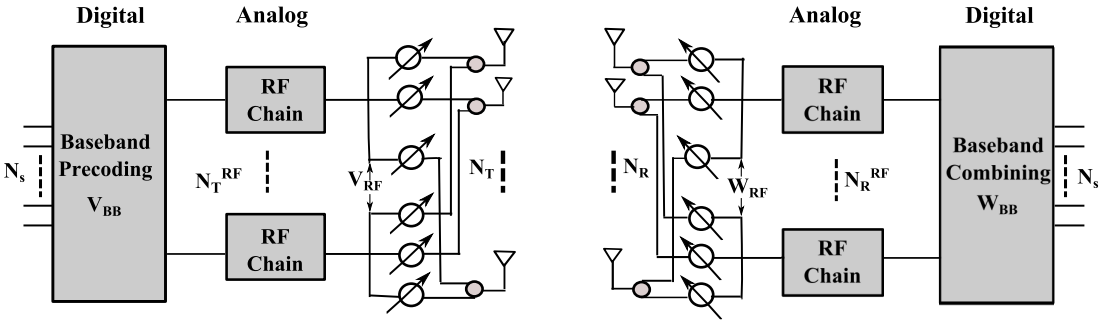


Figure 2.10: Hybrid analog–digital beamforming transceiver architecture.

Hybrid beamforming was proposed to enable increased benefits of MIMO communication at mmWave frequencies; for example, spatial multiplexing of numerous data streams and interference suppression in multi-user transmissions. This architecture splits the processing between RF and baseband domains, as illustrated in Fig. 2.10. This solution accounts for the hardware constraints of mmWave systems and requires a reduced number of RF chains. Herein, the received signal is given by

$$\mathbf{y} = \mathbf{W}_{\mathrm{BB}}^{H}\mathbf{W}_{\mathrm{RF}}^{H}\mathbf{H}\mathbf{V}_{\mathrm{RF}}\mathbf{V}_{\mathrm{BB}}\mathbf{x} + \mathbf{W}_{\mathrm{BB}}^{H}\mathbf{W}_{\mathrm{RF}}^{H}\mathbf{z}, \tag{2.17}$$

where $\mathbf{V}_{\mathrm{RF}}$ denotes the $N_{\mathrm{T}} \times N_{\mathrm{T}}^{\mathrm{RF}}$ RF precoding matrix corresponding to the analog beamforming which uses $N_{\mathrm{s}} \leq N_{\mathrm{T}}^{\mathrm{RF}} \leq N_{\mathrm{T}}$ RF chains and analog phase shifters, while $\mathbf{V}_{\mathrm{BB}}$ is the $N_{\mathrm{T}}^{\mathrm{RF}} \times N_{\mathrm{s}}$ digital precoder used for baseband processing. Similarly, the receiver utilises the $N_{\mathrm{R}} \times N_{\mathrm{R}}^{\mathrm{RF}}$ RF combiner ($\mathbf{W}_{\mathrm{RF}}$) and the $N_{\mathrm{R}}^{\mathrm{RF}} \times N_{\mathrm{s}}$ baseband combiner ($\mathbf{W}_{\mathrm{BB}}$) to process the transmitted signal, where $N_{\mathrm{R}}^{\mathrm{RF}}$ denotes the number of RF chains used by the receiver such that $N_{\mathrm{s}} \leq N_{\mathrm{R}}^{\mathrm{RF}} \leq N_{\mathrm{R}}$.

Hybrid beamforming architectures are categorised into two main structures according to the mapping from RF chains to antennas, namely

- fully-connected structure: in this configuration, each RF chain is connected to all antenna elements via phase shifters; and
- partially-connected structure: here, the entire antenna array is divided into sub-arrays, each of which is connected to one RF chain through the phase shifters.

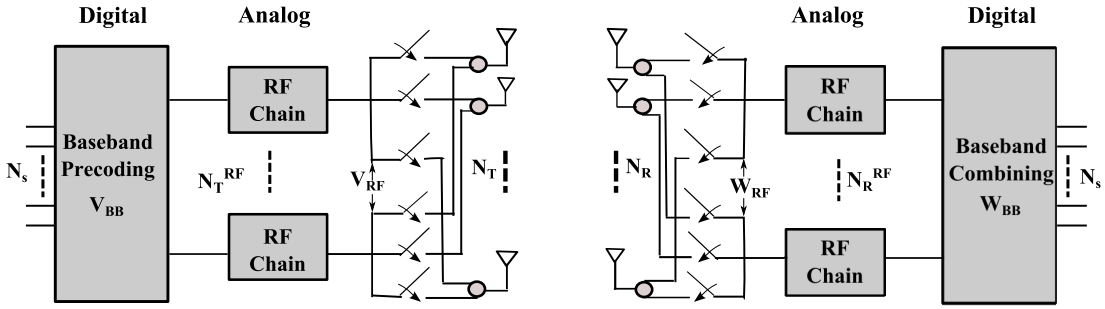


Figure 2.11: Hybrid analog–digital beamforming transceiver architecture with switches.

A detailed comparison of the performance and complexity of both structures is provided in [43], showing that the fully-connected configuration yields higher spectral-efficiency gain since it enables full array flexibility, whereas the partially-connected structure provides higher energy-efficiency since it uses a reduced number of RF chains and reduces the power consumption. Nonetheless, whichever structure is utilised, hybrid beamforming provides a good trade-off between hardware costs, complexity and spectral efficiency. [28, 44, 42].

The RF processing in the hybrid architecture can be implemented with other analog strategies apart from phase shifters, such as switches and lenses. The hybrid solution making use of a switching network has been proposed as providing a further reduction in power consumption and complexity costs over the hybrid architecture using phase shifters [45, 46]. This configuration, illustrated in Fig. 2.11, relies on the sparse nature of mmWave channels to utilise compressed spatial sampling so that the analog combining vectors are designed using only a subset of the antennas rather than overall quantised phases, as in the phase-shifter operation. However, a limitation of the switching network is that as the number of antennas increase, there is a corresponding increase in insertion losses and cross-talk distortion which result in a degradation of the quality of the output signal [47].

The hybrid architecture realised with the lens operation utilises continuous aperture antennas and phased arrays to provide orthogonal beams and enables direct access to the beamspace channel. The continuous aperture phase (CAP) MIMO architecture [48, 49] is one of such architectures which exploits the fact that lenses can compute a spatial Fourier transform which enables the precise control of spatial beams and can be used for link optimization with dramatically lower transceiver costs and complexity. To quantify this, it requires only N_s

RF chains, where N_{s} denotes the number of streams, as opposed to $N_{\mathrm{T}}/N_{\mathrm{R}}$ RF chains required for baseband processing in standard MIMO architectures. Other configurations include the passive and digital multi-beam antenna systems have been shown to improve the spectral and energy efficiencies of 5G networks in [50]. However, there are some anomalies in lenses which impact the their focusing ability. An overview of different lens-types with their corresponding performance quality is presented in [51].

2.3.3 Low-resolution receivers

Typically, in a digital communication system, the receivers are implemented with high-precision analog-to-digital converters (ADCs) to convert the analog baseband signal to digital form. The sampling rate of the ADC is connected to the bandwidth of the system. This implies that high-speed and high-precision ADCs will be required for accurate signal conversion in mmWave systems with large bandwidths. Unfortunately, these high-precision ADCs are power-hungry and expensive for portable devices [52], resulting in a major power bottleneck for implementing mmWave systems. An alternative approach to deal with these prohibitive costs of baseband processing in mmWave systems and address the hardware constraints in mmWave transmissions is to utilise low-precision ADCs in each RF chain at the receiver [52, 53, 54, 55]. In other words, the quantisation resolution of the ADCs is reduced to a few bits (typically 1–5 bits [53]), leading to a reduction of the power consumption of the ADCs.

Fig. 2.12 illustrates the use of 1-bit ADCs in a mmWave receiver to quantise the demodulated signals at the output of each RF chain. These receivers can be implemented at negligible costs and low power consumption since 1-bit ADCs do not need linear amplifiers or automatic gain control [54]. The use of 1-bit ADCs impacts the system design differently from their higher-resolution counterparts. For instance, the capacity-maximising transmit signals are discrete and not continuous. In addition, although the use of 1-bit quantisers generally incurs some capacity loss, it was shown in [55] that, at low signal-to-noise ratios, the MIMO capacity is not severely degraded. In particular, the performance gap between infinite-resolution and 1-bit ADCs is only 1.96 dB while [56] shows that ADCs of $4-5$ bits yield the optimal performance in the presence of hardware impairments. These highlight low-resolution receivers as a promising method to utilise baseband processing at mmWave frequencies.

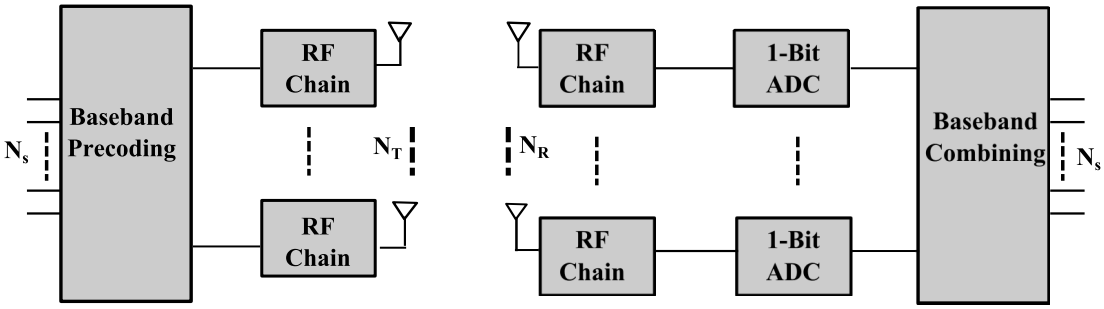


Figure 2.12: An illustration of a 1-bit mmWave receiver.

2.4 Channel Estimation

Channel estimation is important for the practical implementation of any wireless communication system, as the assumption of perfect channel knowledge is highly idealistic. In fact, the available CSI is usually obtained from channel estimates. In mmWave systems, channel estimation is particularly useful for constructing the analog and digital beamformers. The channel estimation strategies utilised in mmWave systems is different from the traditional methods employed to estimate standard wireless channels in a number of respects [57].

- mmWave channel matrices are very large due to the massive antenna arrays characteristic of the propagations. Following traditional techniques would require training a huge number of channel coefficients resulting in a problematic amount of training overhead.

- The entries of a mmWave channel cannot be directly accessed as the measurements of the channel in the baseband is interwoven with the RF precoders and combiners.

- Direct application of traditional estimation techniques will require long training sequences due to the large channel bandwidths available at mmWave, which gives rise to increasing noise power and results in low SNR before constructing the beamformers.

In view of this, we present a brief overview of channel estimation strategies suitable for mmWave communication.

2.4.1 Channel estimation in the analog architecture

Beam training is the main strategy used to configure the analog beamformers in mmWave systems. This method does not require a definite knowledge of the channel state; it utilises a codebook which contains varying beam patterns at different resolutions and relies on an iterative transfer of information between the transmitter and receiver. Specifically, the iterative process involves reducing the beamwidth at each step of information exchange until the best angular directions are obtained, i.e., the combination of angles of departure and arrival that provide the strongest signal between the transmitter and receiver.

Several codebook beam-training protocols have been developed [19, 58, 59]. In [58], the codebooks contain beam patterns classified according to their resolution. The lowest is the quasi-omni antenna pattern which describes a broad region of interest around the devices. The second-level resolution patterns are the sectors which indicate the direction of array patterns covering a relatively broad area of beams (consecutive and non-consecutive), while the highest-resolution patterns of the codebook are the beam patterns. The beamforming training protocol, which aims to configure the beamformers in the minimum time possible, is divided into three stages: (a) device-to-device (D2D) linking, the selection of the best quasi-omni pattern for directional transmission between the devices; (b)sector-level searching, which finds the optimal sector pair in the quasi-omni pattern; and (c) beam-level searching, which obtains the best beam pair in the optimal sectors. In [59], after the optimum beam angle pair has been selected, unique signature codes are assigned to each beam so they can be easily distinguished by the receivers. This results in a minimisation of the training overhead. In [19], the beamforming technique uses a distance-based hierarchical codebook adaptively to sample the channel subspace and search for the beam pair that maximizes the received SNR.

These codebook beam-training strategies are desirable as they eliminate the need for explicit channel estimation. In fact, they have been adopted for implementation in the recent standards for mmWave systems such as IEEE 802.11ad and 1EEE 802.15.3c. However, the information provided by beam training is not sufficient for sophisticated processing applications such as interference cancellation and multi-user MIMO and, as such, these standards assume single-stream transmission [7, 60]. Moreover, the training time involved in beam-training can be prohibitive [59] and the strategies are not robust to identifying NLOS paths

[61].

2.4.2 Channel estimation based on the hybrid architecture

In the hybrid architecture, which combines analog and digital beamforming, the channel estimates are obtained by using compressed sensing (CS) techniques to exploit the inherent sparsity of the mmWave channels. Given that the mmWave channel can be completely described by a small number of parameters, i.e., the angle of arrival (AoA)/angle of departure (AoD) and the complex gain of each path, the estimation of the channel is then equivalent to estimating these key parameters [57].

The basic concept of CS is to recover a compressible (sparse) signal from a limited number of measurements. Specifically, for the hybrid architecture, the hybrid precoders and combiners form the measurement matrices. Following from Fig. 2.10, suppose the transmitter uses G_{T} beamforming vectors to transmit the training symbols and the receiver uses G_{R} measurement vectors to detect the signal transmitted. Then, the $G_{\mathrm{R}} \times G_{\mathrm{T}}$ received signal can be written as

$$\mathbf{Y} = \mathbf{W}_t^H \mathbf{H} \mathbf{V}_t \mathbf{X} + \mathbf{Q}, \tag{2.18}$$

where $\mathbf{W}_t = \mathbf{W}_{\mathrm{RF}} \mathbf{W}_{\mathrm{BB}}$ is the hybrid $N_{\mathrm{R}} \times G_{\mathrm{R}}$ training combining matrix and $\mathbf{V}_t = \mathbf{V}_{\mathrm{RF}} \mathbf{V}_{\mathrm{BB}}$ is the corresponding hybrid $N_{\mathrm{T}} \times G_{\mathrm{T}}$ training precoding matrix. $\mathbf{Q} \in \mathbb{C}^{G_{\mathrm{R}} \times G_{\mathrm{T}}}$ represents the noise matrix and $\mathbf{X}$ is the diagonal matrix containing G_{T} transmitted training symbols.

To exploit the sparsity of the mmWave channel, the received signal matrix in equation (2.18) is vectorized and expressed as [60]

$$\mathbf{y} = \sqrt{P} \left(\mathbf{V}_t^T \otimes \mathbf{W}_t^H \right) \mathbf{A}_D \mathbf{h}_\alpha + \mathbf{z}_Q, \tag{2.19}$$

where equation (2.19) is obtained from applying the virtual channel representation in equation (2.7), with quantised AoAs/AoDs taken from a uniform grid of M points, and assuming that the transmitted symbols are equal, i.e., $\mathbf{X} = \sqrt{P} \mathbf{I}_{G_{\mathrm{T}}}$. $\mathbf{A}_D$ represents the $N_{\mathrm{T}} N_{\mathrm{R}} \times M^2$ dictionary matrix with its columns corresponding to the Kronecker product of the transmitter and receiver array response vectors, $\mathbf{a}_{\mathrm{T}}^H(\hat{\phi}_u) \otimes \mathbf{a}_{\mathrm{R}}(\hat{\theta}_v)$, with $\hat{\phi}_u$ and $\hat{\theta}_v$ being the angles taken from the quantised grid.

Further, $\mathbf{h}_\alpha = \mathrm{vec}\left(\mathbf{H}_\alpha\right)$ denotes the $M^2 \times 1$ vector containing the path gains corresponding to the quantised directions. From equation (2.19), the channel estimation problem can be formulated as [7]

$$\min_{\mathbf{h}_\alpha} \|\mathbf{h}_\alpha\|_0, \tag{2.20a}$$

$$\text{s.t.} \|\mathbf{y} - \sqrt{P}\left(\mathbf{V}_t^T \otimes \mathbf{W}_t^H\right)\mathbf{A}_D\mathbf{h}_\alpha\|_2 \leq \sigma. \tag{2.20b}$$

The problem described by equations (2.20a)–(2.20b) is the generalized channel estimation problem based on hybrid architecture. Additional constraints maybe added depending on the hardware used for analog processing. Specifically, if phase shifters are used, then the RF precoding and combining matrices must have unit norm entries, whereas if switches are used, each column of the RF precoding and combining matrices must have exactly 1 at the index of the selected antenna and 0 elsewhere [46]. Nevertheless, we only have to detect the non-zero elements of $\mathbf{h}_\alpha$ (i.e., the dominant paths of the channel) with the corresponding AoAs and AoDs to solve the channel estimation problem described by equations (2.20a)–(2.20b). Thus, those equations pose a sparse recovery problem which can be efficiently solved with adaptive CS algorithms that define a sensing matrix to recover the non-zero elements of $\mathbf{h}_\alpha$ with high probability and a small number of measurements [60, 62] or with the use of standard greedy recovery algorithms such as the multi-grid orthogonal matching pursuit (OMP) [46].

2.4.3 Channel estimation with 1-bit ADCs

Channel estimation methods for mmWave systems employing the low-resolution receivers also rely on the sparsity of the channel and utilise the narrowband virtual channel model. Therefore, the sparse recovery problem can be formulated.

Considering the system with a 1-bit architecture illustrated in Fig. 2.12, the quantised processed signal is given as

$$\mathbf{y} = \mathrm{sgn}\left(\mathbf{H}\mathbf{s} + \mathbf{z}\right). \tag{2.21}$$

If the transmitter uses G_T beamforming vectors to transmit the training symbols, the received signal obtained by concatenating G_R received vectors is given by [32]

$$\mathbf{Y} = \mathrm{sgn}\left(\mathbf{H}\mathbf{X} + \mathbf{Q}\right), \tag{2.22}$$

where $\mathbf{X} \in \mathbb{C}^{N_\mathrm{T} \times K}$ is the training sequence with K representing the length of the sequence, and $\mathbf{Q}$ is the noise matrix. Using the virtual channel model in equation (2.8) and setting $\mathbf{X} = \mathbf{U}_\mathrm{T}\mathbf{Z}$, we have

$$\begin{aligned} \mathbf{Y} &= \operatorname{sgn}\left(\mathbf{U}_\mathrm{R}\mathbf{H}_\alpha\mathbf{U}_\mathrm{T}^H\mathbf{U}_\mathrm{T}\mathbf{Z} + \mathbf{Q}\right), && (2.23) \\ &= \operatorname{sgn}\left(\mathbf{U}_\mathrm{R}\mathbf{H}_\alpha\mathbf{Z} + \mathbf{Q}\right). && (2.24) \end{aligned}$$

Exploiting the sparse nature of the mmWave channel through vectorization, we obtain

$$\begin{aligned} \operatorname{vec}(\mathbf{Y}) &= \operatorname{sgn}\left(\operatorname{vec}\left(\mathbf{U}_\mathrm{R}\mathbf{H}_\alpha\mathbf{Z} + \mathbf{Q}\right)\right) && (2.25) \\ \mathbf{y}_v &= \operatorname{sgn}\left(\left(\mathbf{Z}^T \otimes \mathbf{U}_\mathrm{R}\right)\mathbf{h}_\alpha + \operatorname{vec}(\mathbf{Q})\right), && (2.26) \end{aligned}$$

where equation (2.26) follows from the identity $\operatorname{vec}(\mathbf{ABC}) = \left(\mathbf{C}^T \otimes \mathbf{A}\right)\operatorname{vec}(\mathbf{B})$, and $\mathbf{h}_\alpha = \operatorname{vec}(\mathbf{H}_\alpha)$. The formulation of $\mathbf{y}_v$ in equation (2.26) represents the sparse recovery formulation for channel estimation using 1-bit ADCs and it involves estimating $\mathbf{h}_\alpha$ given $\mathbf{Z}$. Accordingly, the compressive sensing framework can be applied to recover the sparse vectors. Assuming prior information about the distribution of $\mathbf{h}_\alpha$ is available, the generalised approximate message passing (GAMP) algorithm can be employed to estimate the channel in quick steps [32]. Other algorithms which do not require prior information of $\mathbf{h}_\alpha$, such as expectation maximisation (EM), can be used to detect the non-zero elements of $\mathbf{h}_\alpha$ [63].

Chapter 3

Multi-user Millimeter Wave Cloud Radio Access Networks with Hybrid Beamforming

3.1 Introduction

As mentioned previously in Chapter 2, the increasing prevalence of high-speed, real-time data applications, such as social networking and high-quality wireless video streaming, is driving greater demand for more bandwidth and faster data rates from mobile users. For example, smartphone traffic, which has experienced tremendous increase over the last two decades, is predicted to dominate the global traffic (taking up 99%) by 2022 [1]. While recent years have seen major advances in physical layer techniques to boost spectral efficiency, the spectrum crunch in existing cellular systems presents a fundamental impediment to further capacity increase. Therefore, the current research direction is to exploit under-utilised or unused bands on the frequency spectrum that have not been previously used for cellular communication.

The mmWave frequencies have advanced as promising bands for the realisation of 5G cellular systems because of their potential to achieve the significant goals of the next generation of mobile cellular systems [4] (a general overview of mmWave bands was provided earlier in section 2.1). Recent works have demonstrated the feasibility of mmWave cellular communications. For example, [65] shows that

Work in this chapter has been published in IEEE Systems Journal, December 2018 [64].

wide coverage and high data rates are achievable through the use of trellis coding in constant-envelope orthogonal frequency division multiplexing based mmWave systems. [66] shows how mmWave bands can be used to provide high bandwidth data transfer operations to ensure reliable connectivity especially in congested urban areas. Moreover, [12, 33] and [67] demonstrate how to deploy practical outdoor mmWave communications via real-time measurements at 28-, 38-, 60-, and 70-GHz mmWave bands made in urban environments.

The distinct features of mmWave communication (outlined earlier in section 2.1.1) indicate that propagation at mmWave frequencies is more involved than just changing the carrier frequency. For example, to measure the impact of blockages on mmWave cellular systems, a two-state model (LOS and NLOS) is used [7, 18, 68, 69]. In addition, the work in [14], which studied cellular communication in a dense urban environment, shows that mmWave systems can offer unparalleled increases in capacity for such dense networks.

Network architecture is critical to providing robust networks, which is a key goal for the upcoming 5G networks [70]. Accordingly, CRAN is also receiving heightened interest in academia and industry as the prime architecture for deployment of dense networks to provide increased capacities and efficiencies [71, 72]. Some works focus on a fixed CRAN model, where RRH nodes are fixed at specific locations. For example, [71, 73, 74] show that for MIMO channels gains can be achieved in terms of spectral efficiency, capacity and energy efficiencies, respectively. Other works focus on randomly distributed RRHs. Scenarios with both single cells and multiple cells are considered in [75] and [76]. These studies, demonstrating different performance metrics, attest to the feasibility of CRAN as a distributed antenna network architecture. However, these models have only been studied for standard MIMO communications. Therefore, since the CRAN architecture can deploy dense networks, this chapter presents an analysis of CRAN architecture using mmWave fronthaul and access links.

The main contribution of our work is in the design of a system that couples two potential candidates – mmWave and CRAN – for next-generation cellular communication into multi-user mmWave CRAN systems. In addition, we build a tractable system using stochastic geometry to show the effects of outage probability, average latency and throughput against the signal-to-interference-noise-ratio (SINR) threshold for mmWave CRAN systems[1]. The performance of this system

[1]Note that this work only provides the performance characterisation of the mmWave CRAN system and relies on the hybrid precoder design from the literature cited therein.

is studied for two association schemes. In the first, the RRH or the baseband unit (BBU) with the best channel participates in the transmission where the total power is given to this channel, also called best channel participation (BCP). In the second, the participating RRH is the one closest to the typical UE, also called nearest-neighbour participation (NNP). We characterise this system under key factors including blockages, antenna gain, path losses, and cluster radius. Exact closed-form expressions of outage probability considering LOS and NLOS channels are derived for both scenarios in the noise-limited case (neglecting ICI) while outer bounds of outage probability are also developed for the interference-limited case.

The rest of this chapter is organised as follows. Section 3.2 provides the system model introducing the propagation assumption, SINR model, association scenarios, fronthaul and traffic models. In section 3.3, we present preliminary statistics used in the analysis of the system model. Next, in section 3.4, we derive the closed-form expressions for outage probability and throughput in a noise-limited scenario, while the outer bounds for the interference-limited counterpart is provided in section 3.5. In section 3.6, we provide derivations for calculating the average latency of the system, and we present numerical results validated by simulations in section 3.7. Finally, section 3.8 provides the concluding remarks.

3.2 System Model

We consider the downlink of a multi-user mmWave CRAN system, which consists of a mmWave transmission network, a BBU, multiple mobile UEs and multiple RRHs, as illustrated in Fig. 3.1. The BBU is the central intelligence unit connected to RRHs by high-bandwidth links called “fronthaul links”, whereas the links between the RRHs and UEs are referred to as “access links”. The interface in the fronthaul is standardised as common public radio interface (CPRI) [77]. To meet the strict requirements imposed on the fronthaul by CPRI, the proposed solution is to split the functionality (from baseband to packet processing and radio functions such as amplification, frequency conversion, ADC and digital-to-analogue converter, or DAC, conversion) between the BBU and RRHs. Thus, the BBU performs functions such as baseband and packet processing while the RRHs perform the radio functions and transfer (send/receive) data from BBU to UEs in the system. Although dedicated fibre links are commonly used at the

fronthaul links, recent European projects including 5G PPP [78] and 5G-XHaul [79] propose that mmWave bands can be used for the transmission of fronthaul data because the centralisation in CRAN requires a large consumption of fibre cores which are scarce and expensive to deploy. Therefore, the transmission links of both fronthaul and access in this system operate in the mmWave frequency bands. Each RRH is distributed in a randomly ordered fashion in $\mathcal{R}^2$ using independent homogeneous PPP, Φ_{RRH} and intensity, λ_{RRH}. Similarly, UEs are randomly distributed as PPP and Φ_{UE} with intensity λ_{UE}. The positioning of UEs are not dependent on the location of RRHs; hence, all point processes are independent of each other[2]. The number of antennas on each RRH, UE and the BBU are denoted as N_{r}, N_{UE} and N_{T}, respectively. We assume that there are a number K of UEs in the system such that $N_{\text{r}} \geq K$ or $N_{\text{T}} \geq K$, and that each RRH communicates with several UEs in one instant. Note that throughout this chapter we focus on the access link transmissions between RRHs and UEs in the circular region in Fig. 3.1, which will be referred to as "cluster $\mathcal{G}$".

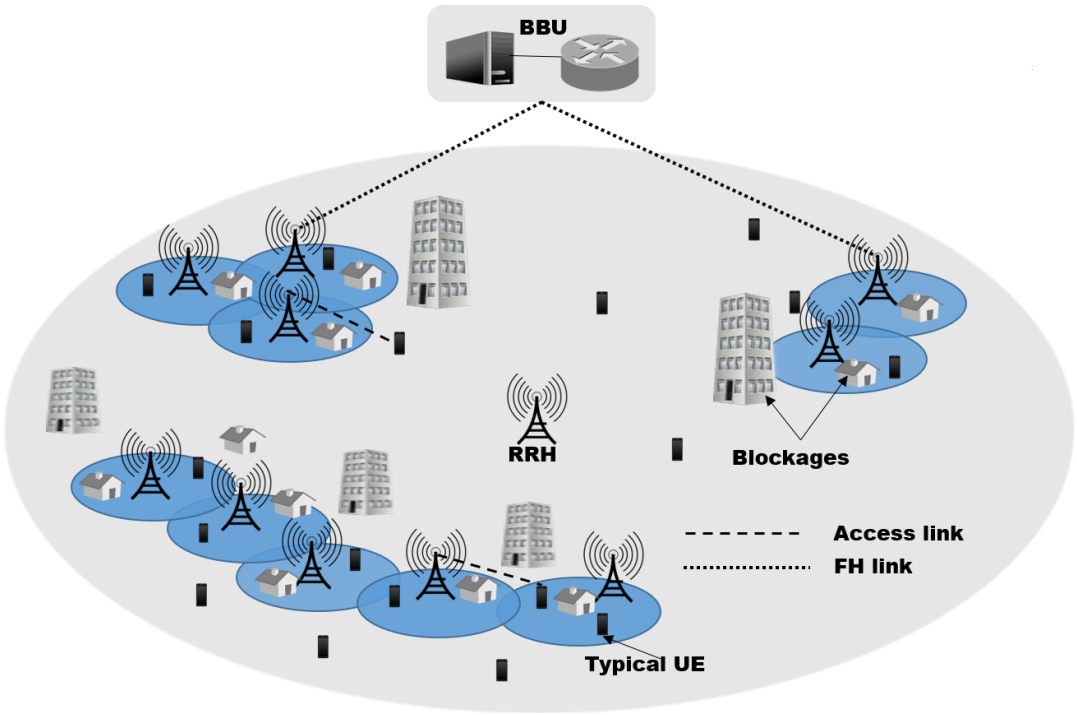


Figure 3.1: An illustration of a multi-user mmWave CRAN system model.

[2]Although the positioning of the UEs are independent of the position of the RRHs, we assume that since the BBU makes the decisions about which RRH serves a user, thus ensuring that the RRHs that serve the UEs are not violating the effective isotropic radiated power (EIRP) limitations.

3.2.1 Propagation assumptions

3.2.1.1 Transmission model

The propagation from any RRH to each UE is via a fully connected hybrid precoder that combines RF and BB precoding. We assume that every UE is served via only one stream, thus it is sufficient that each UE employs an RF-only combiner to decode the transmitted signal, as described in [80]. Hybrid precoding is also used to cancel out the unwanted signals of other UEs. Considering a typical UE located at the origin of the cluster, UE_0, the clustered mmWave channel between the kth RRH located at position y in $\mathcal{G}$ and UE_0 is given as

$$\mathbf{H}_{k,0} = \sqrt{\frac{N_{\mathrm{r}} N_{\mathrm{UE}}}{L(k,0)}} \sum_{u=1}^{W_{k,0}} \eta_{u,k,0}\, \mathbf{a}_{\mathrm{UE}}(\theta_{u,k,0}) \mathbf{a}^{\mathrm{H}}_{\mathrm{RRH}}(\phi_{u,k,0}), \tag{3.1}$$

where $\eta_{k,0}$ is the complex gain, assumed to follow a normal distribution with zero mean and unit variance for both LOS and NLOS links to enhance analytical tractability [42, 69], θ is the AoA, ϕ is the AoD, $W_{k,0}$ is the number of paths from the kth RRH to the UE_0,[3] $L(k,0)$ is the path loss given as $L(k,0) = r_{k,0}^{\alpha_i}$, with $i \in \mathrm{L}, \mathrm{N}$. Here α_i denotes the path-loss exponent and can either be LOS and NLOS depending on the link between them, and $\mathbf{a}_{\mathrm{RRH}}$ and $\mathbf{a}_{\mathrm{UE}}$ are the array response vectors of each RRH and UE, respectively.

Due to the sparsity of mmWave channels, we assume that all scattering happens in the azimuth plane, and model these array vectors as ULAs. Hence, the array response vector at the RRH is given by

$$\mathbf{a}_{\mathrm{RRH}}(\phi) = \frac{1}{\sqrt{N_{\mathrm{r}}}} \left[1, e^{i\frac{2\pi}{\lambda} d \sin(\phi)}, \ldots, e^{(N_{\mathrm{r}}-1) i \frac{2\pi}{\lambda} d \sin(\phi)}\right], \tag{3.2}$$

where λ is the wavelength and d is the distance between antenna elements. The array response vectors at the UE ($\mathbf{a}_{\mathrm{UE}}(\theta)$) and BBU ($\mathbf{a}_{\mathrm{BBU}}(\phi)$) are written in similar fashion.

3.2.1.2 Received signal

We denote $\mathcal{U}$ as the maximum number of UEs in $\mathcal{G}$ to which a single RRH can instantaneously communicate[4], and assume that the maximum value of $\mathcal{U}$ is the

[3] Unit paths are assumed for each UE in this system throughout this chapter.

[4] This assumption is to ensure that an RRH is not overloaded.

number of RF chains ($\mathcal{U} \leq N_{\mathrm{RF}}$) since the hybrid precoding gain is constrained by $\min(N_{\mathrm{RF}}, K)$ [81]. For the links between the kth RRH in cluster $\mathcal{G}$ and the $\mathcal{U}$ UEs connected to it, the BB and RF precoders are presented, respectively, as

$$\mathbf{V}_{\mathrm{BB}}^{k} = \left[\mathbf{v}_{\mathrm{BB}}^{k,1}, \mathbf{v}_{\mathrm{BB}}^{k,2}, \ldots, \mathbf{v}_{\mathrm{BB}}^{k,\mathcal{U}}\right] \tag{3.3}$$

$$\mathbf{V}_{\mathrm{RF}}^{k} = \left[\mathbf{v}_{\mathrm{RF}}^{k,1}, \mathbf{v}_{\mathrm{RF}}^{k,2}, \ldots, \mathbf{v}_{\mathrm{RF}}^{k,\mathcal{U}}\right]. \tag{3.4}$$

At the typical UE, the RF combiner $\mathbf{w}_{\mathrm{RF}}^{0}$ is used to process the received signal. The processed signal is given by

$$y_0 = \varrho_{k,0}\mathbf{h}_{k,0}^{\mathrm{eff}}\mathbf{v}_{\mathrm{BB}}^{k,0}s_0 + \sum_{\mathrm{g}\in\mathcal{U},\mathrm{g}\neq 0}\varrho_{k,\mathrm{g}}\mathbf{h}_{k,\mathrm{g}}^{\mathrm{eff}}\mathbf{v}_{\mathrm{BB}}^{k,\mathrm{g}}s_{\mathrm{g}} + z_0, \tag{3.5}$$

where s_0 denotes the transmitted symbol for the typical UE, while s_g denotes the transmitted symbol for a user other than the typical UE associated with RRH_k, $\mathbf{h}_{k,0}^{\mathrm{eff}} = (\mathbf{w}_{\mathrm{RF}}^{0})^{\mathrm{H}}\mathbf{H}_{k,0}\mathbf{V}_{\mathrm{RF}}^{k}$ is a $1 \times N_{\mathrm{RF}}$ vector called the "effective channel", $\varrho_{k,0}$ is the average received power given as $\varrho_{k,0} = \frac{P_T}{K}$, z_0 is additive white Gaussian noise (AGWN) such that $z_0 \sim \mathcal{CN}(0, \sigma_z^2)$ and P_T is the total transmit power enforced by the normalization of V_{BB} such that $||\mathbf{V}_{\mathrm{RF}}\mathbf{V}_{\mathrm{BB}}||_F^2 = N_{\mathrm{s}}$ with N_{s} representing the number of data streams [28].

To design the hybrid beamformers, we adopt the two-stage hybrid scheme proposed in [80][5]. Here, in the first stage, the kth RRH and the typical UE connected to it design the RF precoder and combiner simultaneously using phase shifters to maximize the desired signal for the typical UE. In the second stage, the RRH designs the baseband precoder for the typical UE to mitigate the interference from other UEs.

Accordingly, the optimization problem providing the RF precoder and combiner from the kth RRH to UE_0 takes the form

$$\mathcal{P}_1 : \left((\mathbf{w}_{\mathrm{RF}}^{0})^{\mathrm{opt}}, (\mathbf{v}_{\mathrm{RF}}^{k,0})^{\mathrm{opt}}\right) = \underset{\mathbf{w}_{\mathrm{RF}}^{0}, \mathbf{v}_{\mathrm{RF}}^{k,0}}{\arg\max} \; |(\mathbf{w}_{\mathrm{RF}}^{0})^{H}\mathbf{H}_{k,0}\mathbf{v}_{\mathrm{RF}}^{k,0}|. \tag{3.6}$$

To solve the problem described by $\mathcal{P}_1$, first we apply the singular value decomposition (SVD) to the channel matrix $\mathbf{H}_{k,0}$, i.e., we set $\mathbf{H}_{k,0} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^{H}$, where $\mathbf{U}$ and $\mathbf{V}$ are $N_{\mathrm{UE}} \times N_{\mathrm{UE}}$ and $N_{\mathrm{RRH}} \times N_{\mathrm{RRH}}$ matrices containing the eigenvectors

[5]This hybrid beamforming design is less computationally complex than other existing solutions to the hybrid beamforming problem in literature such as the manifold optimisation method [82] and provides results close to that of the unconstrained beamformers.

of $\mathbf{H}_{k,0}(\mathbf{H}_{k,0})^H$ and $(\mathbf{H}_{k,0})^H\mathbf{H}_{k,0}$, respectively, while $\mathbf{\Sigma}$ is an $N_{\mathrm{UE}} \times N_{\mathrm{RRH}}$ diagonal matrix consisting of the eigenvectors of $\mathbf{H}_{k,0}(\mathbf{H}_{k,0})^H$ in decreasing order.

Lemma 3.1. *[83, Theorem 3]. For a large number of antennas, the optimal RF precoder and combiner* $\left((\mathbf{w}_{\mathrm{RF}}^0)^{opt}, (\mathbf{v}_{\mathrm{RF}}^{k,0})^{opt}\right)$ *are obtained when right and left singular vectors corresponding to the non-zero eigenvalues of* $\mathbf{H}_{k,0}$ *converge in chordal distance to* $\boldsymbol{a}_{\mathrm{UE}}(\theta_{u,k,0})$ *and* $\boldsymbol{a}_{\mathrm{RRH}}(\phi_{u,k,0}) \,\forall\, u = 1, \ldots, W_{k,0}$ *when* $W_{k,0} \ll \min(N_{\mathrm{RRH}}, N_{\mathrm{UE}})$.

From the above lemma, $\mathbf{w}_{\mathrm{RF}}^0 = \mathbf{a}_{\mathrm{UE}}(\theta_{u_{\max},k,0})$ and $\mathbf{v}_{\mathrm{RF}}^{k,0} = \mathbf{a}_{\mathrm{RRH}}(\phi_{u_{\max},k,0})$, where $u_{\max} = \arg\max_u |\eta_{u,k,0}|$ indicates the path with the maximum gain.

Next, the BB precoders are designed such that interference from other UEs in $\mathcal{G}$ is cancelled. Within this context, a simple ZF precoder can be utilised. Accordingly, the BB precoder for the typical UE from kth RRH is given as

$$\mathbf{v}_{\mathrm{BB}}^{k,0} = (\mathbf{h}_{k,0}^{\mathrm{eff}})^{\mathrm{H}} \left(\mathbf{h}_{k,0}^{\mathrm{eff}} (\mathbf{h}_{k,0}^{\mathrm{eff}})^{\mathrm{H}} \right)^{-1}. \tag{3.7}$$

3.2.1.3 Blockage model

To model blockages in this mmWave network, we use the probabilistic model validated in [18] which defines a link of length a as LOS with probability p_{L},[6] if a is less than or equal to the radius of some region D; $a \leq D$.[7] However, when $a > D$, the link is NLOS with probability of p_{N}.

3.2.2 SINR model

To model the SINR of a mmWave network, we consider the two different schools of thought about propagation in mmWave networks; on one hand, the authors of [12, 18, 84] mention that mmWave networks tend to be noise-limited due to high blockage density, which make signals from unwanted sources negligible. On the other hand, the authors of [11, 22, 42] consider high base station density with moderate blockages and assume that mmWave networks tend to be interference-limited. Therefore, we model the SINR in both noise-limited and interference-limited cases.

[6] It is important to note that the values of p_L are dependent on the geography of an area i.e. a low value is assumed for dense urban areas and a higher one for semi-urban areas.

[7] D is defined in [18] as the radius of a circle from actual measurements in urban regions of New York

3.2.2.1 Noise-limited case

In this scenario, we consider only intra-cluster interference from other users associated with an RRH. The analysis is performed for a typical UE following Slivnyak's theorem. The SINR between the kth RRH and the typical UE is obtained from equation (3.5). Denoting SINR as ρ, we obtain

$$\rho_{k,0} = \frac{\varrho_{k,0} \|\mathbf{h}_{k,0}^{\text{eff}} \mathbf{v}_{\text{BB}}^{k,0}\|^2}{\sum_{\substack{\text{g}\in \text{n}_\text{y} \\ \text{g}\neq 0}} \varrho_{k,\text{g}} \|\mathbf{h}_{k,\text{g}}^{\text{eff}} \mathbf{v}_{\text{BB}}^{k,\text{g}}\|^2 + \sigma_z^2}, \tag{3.8}$$

where the first term of the denominator denotes the intra-cluster interference to the typical UE, which is zero after successful interference cancellation, and σ_z^2 is the noise variance.

For tractability in the statistical analysis in this paper, we rewrite the SINR in equation (3.8) as

$$\rho_{k,0} \approx \frac{\text{B}_{k,0}\, \eta_{k,0}^2\, \omega_p(N_\text{r}, \mathcal{U})}{r_{k,0}^{\alpha_i}\, \sigma_z^2}, \tag{3.9}$$

where $\text{B}_{k,0} = \varrho_{k,0}\, N_\text{r}\, N_\text{UE}$, and $\omega_p(N_\text{r}, \mathcal{U})$ is the precoding penalty defined as

$$\omega_p(N_\text{r}, \mathcal{U}) = \begin{cases} 1, & \text{w.p. } \left(1 - \frac{1}{N_\text{r}}\right)^{\mathcal{U}-1} \\ 0, & \text{otherwise.} \end{cases} \tag{3.10}$$

The explanation of this penalty can be found in [69, Proposition 1].

3.2.2.2 Interference-limited case

In this scenario, the typical UE experiences interference from within the cluster as well as from neighbouring clusters. It should be noted that the RRHs from other clusters causing interference are those minimally affected by blockages. Consequently, the received signal after applying the RF precoders and combiners to the transmitted signal is given as

$$y_0 = \varrho_{k,0} \mathbf{h}_{k,0}^{\text{eff}} \mathbf{v}_{\text{BB}}^{k,0} s_0 + \sum_{\text{g}\in\mathcal{U}, \text{g}\neq 0} \varrho_{k,\text{g}} \mathbf{h}_{k,\text{g}}^{\text{eff}} \mathbf{v}_{\text{BB}}^{k,\text{g}} s_\text{g} + \mathbf{I}_{\Phi'_{\text{BS}}} + z_0 \tag{3.11}$$

where s_0 and s_g denotes the transmitted symbols for the typical UE and for another UE associated with RRH_k, $\mathbf{I}_{\Phi'_{\text{BS}}}$ is the inter-cluster interference, and $\mathbf{z}_0$

is the complex Gaussian noise, $\mathbf{z}_0 \sim \mathcal{CN}(0, \sigma_z^2)$.

Thus, the SINR of the typical UE served by the kth RRH is obtained from the received signal in equation (3.11) and expressed as

$$\mathrm{SINR_{k,0}} = \frac{\varrho_{k,0} \left|\mathbf{h}_{k,0}^{\mathrm{eff}} \mathbf{v}_{\mathrm{BB}}^{0}\right|^2}{\sigma_z^2 + \sum\limits_{\mathrm{g}\in\Phi'_{\mathrm{UE}},\mathrm{g}\neq 0} \varrho_{k,g} \left|\mathbf{h}_{k,g}^{\mathrm{eff}} \mathbf{v}_{\mathrm{BB}}^{g}\right|^2 + \mathbf{I}_{\Phi'_{\mathrm{BS}}}}, \tag{3.12}$$

where $\varrho_b = \frac{\mathrm{P_{BS}}}{\mathcal{U}_b}$ with $\mathcal{U}_b$ denoting the number of users connected to a base station (BS) in another cluster, and $\Phi'_{\mathrm{BS}} = \Phi_{\mathrm{BS}} \setminus \mathcal{G} \cap \Phi_{\mathrm{BS}}$.

Next, the SINR in equation (3.12) is approximated for the statistical analysis and expressed as:

$$\mathrm{SINR_{k,0}} \approx \frac{G_0 \left|\eta_{\mathrm{k,0}}\right|^2 \mathrm{r}_{k,0}^{-\alpha_i}\, \omega_p(N_{\mathrm{r}}, \mathcal{U})}{\sigma_z^2 + \mathbf{I}_{\Phi'_{\mathrm{BS}}}}, \tag{3.13}$$

where $G_0 = \varrho_{k,0}\, N_{\mathrm{UE}} N_{\mathrm{r}}, \eta_{k,0} = \mathbf{h}_{k,0}^{\mathrm{eff}} \mathbf{v}_{\mathrm{BB}}^{0}$, $r_{k,0}$ is the distance between the typical UE and the kth RRH, α_i is the path-loss exponent which could be LOS or NLOS,[8] $\omega_p(N_{\mathrm{r}}, \mathcal{U})$ is the precoding penalty similarly defined in equation (3.10) and $\mathbf{I}_{\Phi'_{\mathrm{BS}}}$ is the inter-cluster interference given as [42]

$$\mathbf{I}_{\Phi'_{\mathrm{BS}}} = \sum_{b\in\Phi'_{\mathrm{BS}},v\neq 0} G_{\mathrm{v}} \left|\eta_{\mathrm{b,v}}\right|^2 \mathrm{r}_{\mathrm{v}}^{-\alpha} \sum_{v\in\mathcal{U}_v} \left|\left|\mathbf{a}_{\mathrm{UE}}(\theta_{v,y,x}) \mathbf{a}_{\mathrm{RRH}}^{\mathrm{H}}(\phi_{v,y,x})\right|\right|^2, \tag{3.14}$$

$$= \sum_{b\in\Phi'_{\mathrm{BS}},v\neq 0} G_{\mathrm{v}} \left|\eta_{\mathrm{b,v}}\right|^2 \mathrm{r}_{\mathrm{v}}^{-\alpha} \Upsilon_v, \tag{3.15}$$

where G_{v} is the gain from another cluster, similarly defined as G_0 but scaled by the precoding penalty.[9], Υ_v is defined as the inner product of the transmit and receive beam steering vectors, expressed as

$$\Upsilon_v \triangleq \begin{cases} 1, & \theta_{v,y,x} = \phi_{v,y,x} \\ \tau_{\mathrm{RRH}}, & \text{otherwise,} \end{cases} \tag{3.16}$$

where $\tau_{\mathrm{RRH}} < 1$ denotes the side lobe gain.

Remark 3.1. *Note that although the analyses in this chapter focus on the downlink transmission, the results are applicable to uplink transmissions, with the roles*

[8] The parameters for α used in simulations are adopted from measurements validated in [10, 18, 14].

[9] For simplicity we assume the sum of penalties from RRHs in interfering clusters to be constant.

of the UEs and RRHs reversed. More specifically, where multiple UEs transmit independent messages to their associated RRHs, the received signal can be formulated similar to equation (3.11)*. In this case, ICI is caused by the other UEs associated with the same RRH as well as those outwith the cluster, and these interferences can be characterised with similar stochastic geometry tools.*

3.2.3 Association scenarios

When a UE associates itself with an RRH and makes a request, the RRHs help the signals from the UE to be decoded by the BBU. Thus, a UE is serviced through the access link (from UE to RRH) and fronthaul link (from tagged RRH to BBU). To analyse the performance of this system, we consider two schemes based on how the UE associates with an RRH in cluster $\mathcal{G}$, namely:

- BCP: in this scenario, BBU determines which RRH has the best channel in cluster $\mathcal{G}$ for transmission to UEs, and
- NNP: this scenario is considered for its ability to reduce overhead whilst achieving acceptable performance; the metric is based on distance where the RRH closest to an UE is selected for transmission.

3.2.4 Traffic model

To measure the quality of service (QoS) and model the traffic delivery from the BBU to an UE via its associated RRH in this mmWave CRAN system, we adopt the queueing theory of [85], denoting the typical UE as UE_0 and its associated RRH as RRH_k. We assume that the traffic arrival times to UEs in the network follows a PPP with parameter λ which denotes the arrival rate per unit area, with an average time of μ seconds.[10] On the account of a constant fronthaul rate, it is reasonable to assume that the traffic delivery time of the fronthaul is also an exponential distribution. Thus, the traffic delivery in fronthaul actualises an M/M/1 queuing system [85].

Similarly, in the access links, the traffic arrival to UE_0 follows a PPP with an average arrival rate of λ_0. However, since the UEs at various locations associated with RRH_k have different rates depending on the channel conditions, the service

[10]This assumption is based on the fact that the UEs are distributed according to a PPP, Φ_{UE}.

time in RRH_k is generally distributed [86]. Therefore, the traffic delivery from RRH_k to UE_0 realises an M/G/1 queueing model [85]. Accordingly, the traffic delivery from BBU to UE_0 via RRH_k is modelled as a combined queue system.

3.2.5 Fronthaul model

For completeness, and to give a general overview of the functionality between the BBU and RRHs in the network, we include the fronthaul model.

Note that for the implementation of CRAN, the BBU does not have to be physical, implying that the network operator can dynamically map radio signals from the BBU to any RRH using software-defined networking (SDN) concepts[87]. As highlighted earlier in section 3.2, CPRI imposes strict requirements on the fronthaul network [77] which determine the functionality between the BBU and RRHs.

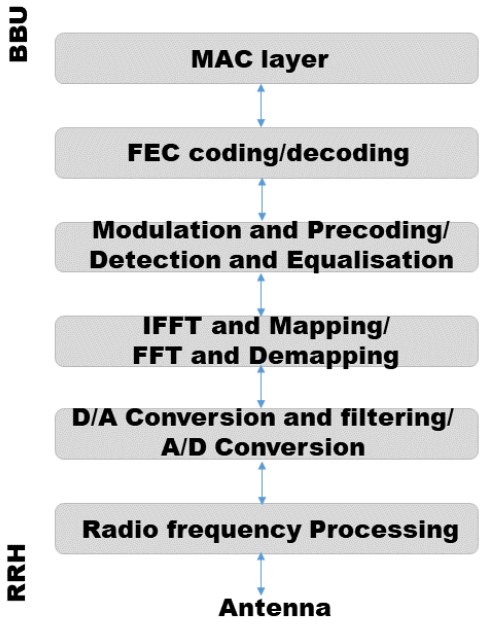


Figure 3.2: CRAN fronthaul-link logical structure [81].

Therefore, Fig. 3.2 presents a logical structure illustrating the downlink and uplink processing chain of the fronthaul link. In the downlink, data packets for the UEs are processed at the media access control (MAC) layer, where appropriate headers and schedules are added on followed by forward error correction (FEC)

encoding and then modulation and precoding. These operations are performed with CSI available at the BBU[11]. Next, an inverse Fourier transform is performed to map UE data to physical resources in time slots and sub-carriers. After that, the data is converted to the analogue domain using a DAC and then up-converted to the carrier frequency and transmitted to the antenna of the RRH. More details of the processing can be found in [88]. Note that for the uplink, the process is reversed as indicated by the upward flow of Fig. 3.2.

3.3 Preliminary Statistics

This section gives preliminary statistics introducing the statistical properties of the mmWave channel and the performance metrics which will be used in the subsequent analysis of the access links.

3.3.1 Millimeter wave channel statistics

If we assume Gaussian symbols are transmitted, then the probability distribution function (PDF) of the signal-to-noise-ratio (SNR), $\rho_{k,0}$, is F-scaled [89] and can be written as

$$f_{\rho_{k,0}}(z) = \left(1 - \frac{1}{N_{\mathrm{r}}}\right)^{\mathcal{U}-1} \frac{N_{\mathrm{r}}!\,(B_{k,o})^{K} z^{N_{\mathrm{r}}-K}}{(K-1)!(N_{\mathrm{r}}-K)!(B_{k,0}+z)^{N_{\mathrm{r}}+1}}. \tag{3.17}$$

The work in [90] shows that the F-distribution can be approximated to chi-square distribution. Therefore, rewriting equation (3.17) in terms of chi-squared distribution with $2(N_{\mathrm{r}} - K + 1)$ degrees of freedom, we obtain

$$f_{\rho_{k,0}}(z) \approx \left(1 - \frac{1}{N_{\mathrm{r}}}\right)^{\mathcal{U}-1} \sum_{i \in \mathrm{L,N}} \frac{(K r_{k,0}^{\alpha_i})^{N_{\mathrm{r}}-K+1}}{(N_{\mathrm{r}}-K)!} \frac{1}{(B_{k,0})^{N_{\mathrm{r}}-K+1}} z^{N_{\mathrm{r}}-K} e^{-\frac{z\,K\,r_{k,0}^{\alpha_i}}{B_{k,0}}}. \tag{3.18}$$

Note that for ease of exposition and without abuse of notation, we will treat the subsequent derivations from (3.18) as exact expressions. Moreover, integrating equation (3.18) for a given distance, $r_{k,0}$, yields a conditional cumulative distri-

[11]The assumption of full CSI availablility is ideal. Notwithstanding, the analyses presented in this work provide an upper bound on the performance of the mmWave CRAN system and can be easily extended to imperfect CSI considerations.

bution function ($F_{\rho_{k,0}|r_{k,0}}$), which is expressed as

$$F_{\rho_{k,0}|r_{k,0}}(z) = \left(1 - \frac{1}{N_{\mathrm{r}}}\right)^{\mathcal{U}-1} \sum_{i \in \mathrm{L,N}} p_i \, \frac{1}{(N_{\mathrm{r}} - K)!} \gamma\left(N_{\mathrm{r}} - K + 1, \frac{z\, K\, r_{k,0}^{\alpha_i}}{B_{k,0}}\right), \quad (3.19)$$

where γ is the lower incomplete gamma function.

3.3.2 Performance metrics

The performance of the network will be measured in terms of outage probability, throughput and latency.

3.3.2.1 Outage probability of a generic RRH

The outage probability of any RRH in cluster $\mathcal{G}$ is given as

$$\mathbb{P}_{out}(\xi) = \int_0^R f_r(r) F_{\rho|r}(\xi)\, \mathrm{d}r, \quad (3.20)$$

where ξ denotes the SNR threshold, $f_r(r)$ represents the uniform distribution of all RRHs in $\mathcal{G}$ given as $f_r(r) = \dfrac{2\,r}{R^2}$ and $F_{\rho|r}$ is the CCDF given as

$$F_{\rho|r}(\xi) = \left(1 - \frac{1}{N_{\mathrm{r}}}\right)^{\mathcal{U}-1} \sum_{i \in \mathrm{L,N}} p_i \, \frac{1}{(N_{\mathrm{r}} - K)!} \gamma\left(N_{\mathrm{r}} - K + 1, \frac{\xi\, K\, r^{\alpha_i}}{B}\right). \quad (3.21)$$

It is worthwhile to notice that the cumulative distribution function (CDF) of received SNR at kth RRH from a BBU follows from equation (3.19) and is expressed as

$$F_{\rho_f}(\xi) = \left(1 - \frac{1}{N_{\mathrm{t}}}\right)^{\mathcal{M}-1} \sum_{i \in \mathrm{L,N}} \frac{p_i\, \gamma\left(N_{\mathrm{t}} - M + 1, \frac{\xi\, M\, r_f^{\alpha_i}}{B_f}\right)}{(N_{\mathrm{t}} - M)!}. \quad (3.22)$$

In the following proposition, we show the overall outage probability of the k^{th} RRH which will be used in analysis of the association schemes.

Proposition 3.1. *The outage probability of the* k^{th} *RRH to a typical UE is given*

as

$$\mathbb{P}_{out}(t) = \left(1 - \frac{1}{N_{\mathrm{r}}}\right) \sum_{i \in \mathrm{L,N}}^{\mathcal{U}-1} p_i \, \frac{2\,(t\,K)^{-\frac{2}{\alpha_i}}}{R^2} \left(1 - \sum_{m=0}^{N_{\mathrm{r}}-K} \frac{\left[\Gamma\left(\frac{2}{\alpha_i} + m\right) - \Gamma\left(\frac{2}{\alpha_i} + m, KR^{\alpha_i} t\right)\right]}{m!}\right). \tag{3.23}$$

Proof. The outage probability of the kth RRH is defined as

$$\mathbb{P}_{out}(z) = \int_0^R f_{r_{k,0}}(r) F_{\rho_{k,0}|r_{k,0}}(z)\, \mathrm{d}r. \tag{3.24}$$

Substituting the expressions of $f_{r_{k,0}}(r) = \frac{2r}{R^2}$ and $F_{\rho_{k,1}|r_k}$ from equation (3.19), we obtain

$$\mathbb{P}_{out}(t) = \int_0^R \frac{2r}{R^2} \left(1 - \frac{1}{N_{\mathrm{r}}}\right) \sum_{i \in \mathrm{L,N}}^{\mathcal{U}-1} p_i \, \frac{1}{(N_{\mathrm{r}} - K)!} \gamma\left(N_{\mathrm{r}} - K + 1, t\,K\,r_{k,0}^{\alpha_i}\right) \mathrm{d}r, \tag{3.25}$$

where $t = \dfrac{\xi}{B_{k,0}}$.

Starting with the LOS link, we have

$$\begin{aligned}
\mathbb{P}_{out}^{\mathrm{L}}(t) &= \int_0^R \left(1 - \frac{1}{N_{\mathrm{r}}}\right)^{\mathcal{U}-1} \frac{2r}{R^2} \frac{p_{\mathrm{L}} \gamma\left(N_{\mathrm{r}} - K + 1, tKr_{k,0}^{\alpha_{\mathrm{L}}}\right)}{(N_{\mathrm{r}} - K)!}\, \mathrm{d}r, \\
&\overset{(a)}{=} \int_0^R \frac{2r}{R^2} \left(1 - \frac{1}{N_{\mathrm{r}}}\right) p_{\mathrm{L}}^{\mathcal{U}-1} \left[1 - \sum_{m=0}^{N_{\mathrm{r}}-K} \frac{(tKr^{\alpha_{\mathrm{L}}})^m}{m!} e^{tKr^{\alpha_{\mathrm{L}}}}\right] \mathrm{d}r, \\
&\overset{(b)}{=} \left(1 - \frac{1}{N_{\mathrm{r}}}\right)^{\mathcal{U}-1} p_{\mathrm{L}} \left[1 - \sum_{m=0}^{N_{\mathrm{r}}-K} \frac{2\,(t\,K)^{-\frac{2}{\alpha_{\mathrm{L}}}}}{R^2} \frac{\left[\Gamma\left(\frac{2}{\alpha_{\mathrm{L}}} + m\right) - \Gamma\left(\frac{2}{\alpha_{\mathrm{L}}} + m, KR^{\alpha_{\mathrm{L}}} t\right)\right]}{m!}\right],
\end{aligned} \tag{3.26}$$

where (a) follows from the series equivalent of the lower Gamma incomplete function given by $\gamma(N,t) = (N-1)!\left(1 - e^{-t}\sum_{k=0}^{N-1} \frac{t^k}{k!}\right)$, and (b) follows from solving the integral with respect to r.

The outage probability of the NLOS link can be derived by following similar steps to obtain

$$\mathbb{P}_{out}^{\mathrm{N}}(t) = \left(1 - \frac{1}{N_{\mathrm{r}}}\right)^{\mathcal{U}-1} p_{\mathrm{N}} \left[1 - \sum_{m=0}^{N_{\mathrm{r}}-K} \frac{2\,(t\,K)^{-\frac{2}{\alpha_{\mathrm{N}}}}}{R^2} \frac{\left[\Gamma\left(\frac{2}{\alpha_{\mathrm{N}}} + m\right) - \Gamma\left(\frac{2}{\alpha_{\mathrm{N}}} + m, KR^{\alpha_{\mathrm{N}}} t\right)\right]}{m!}\right]. \tag{3.27}$$

Finally, the proof of (3.23) can be concluded by the summation of $\mathbb{P}_{out}^{\mathrm{L}}(t)$ and $\mathbb{P}_{out}^{\mathrm{N}}(t)$.

3.3.2.2 Throughput

By definition, throughput $\mathcal{R}$ is given by

$$\mathcal{R} = \log_2(1+\rho)\left(1-\mathbb{P}_{out}\right), \tag{3.28}$$

where ρ is the received SINR. The throughput can be obtained by substituting $\mathbb{P}_{\mathrm{out}}$ from equation (3.23) with the appropriate values of ρ for the scenario under consideration.

3.3.2.3 Latency ratio

This performance metric is used to measure the QoS of the network. It is defined as the delayed UE duration per unit of service time during the transmission process and will be discussed in detail in section 3.6.

3.4 Performance in a Noise-Limited Scenario

In this section, we study the access link outage probability performance of our system model in a noise-limited scenario under the two different user-association scenarios aforementioned in section 3.2.3.

3.4.1 NNP

Outage occurs in this scenario when the channel of an RRH closest to the typical UE is in outage. Note that throughout this section, we denote the kth RRH as the RRH closest to the typical user. Additionally, for this scenario, the BBU transmits directly to the user only when there are no RRHs in $\mathcal{G}$. Therefore, the outage probability for the nearest neighbour is defined by the following proposition:

Proposition 3.2. *The outage probability of an RRH closest to the typical* UE *is*

expressed as

$$\begin{aligned}\mathbb{P}_{out}^{\mathrm{NNP}}(\xi) &= \left(1-\frac{1}{N_{\mathrm{t}}}\right)^{\mathcal{U}-1}\sum_{i\in\mathrm{L,N}}\frac{p_i\ \gamma\left(N_{\mathrm{t}}-K+1,\,\xi\,K\,R^{\alpha_i}\right)}{(N_{\mathrm{t}}-K)!}\\ &\times\left(1-\frac{1}{N_{\mathrm{r}}}\right)^{\mathcal{U}-1}\frac{e^{-\lambda_{\mathrm{RRH}}\,\pi\,R^2}}{(N_{\mathrm{r}}-K)!}\sum_{n=0}^{\infty}\frac{(-1)^n}{n!}\frac{(\xi\,K)^{(N_r-K+n+1)}}{(N_r-K+n+1)}\\ &\times\left[\sum_{i\in L,N}p_i\,\frac{\gamma\frac{(\alpha_i(N_r-K+n+1)}{2}+1,\pi\lambda_{\mathrm{RRH}}R^2)}{(\pi\lambda_{\mathrm{RRH}})^{\frac{(\alpha_i(N_r-K+n+1)}{2}}}\right].\end{aligned}\tag{3.29}$$

Proof. During the NNP scenario, the typical UE is served by the RRH closest to it. Denoting r as the distance between the typical UE and its closest RRH, we can express its PDF as [91]

$$f_{\mathrm{closest}} = 2\,\pi\,\lambda_{\mathrm{RRH}}\,r\,e^{-\lambda_{\mathrm{RRH}}\,\pi\,r^2}.\tag{3.30}$$

Consequently, the outage probability for the closest RRH to the typical UE is given by

$$\mathbb{P}_{out}^{\mathrm{NNP}}(\xi) = \int F_{\rho|r}(\xi)f_{\mathrm{closest}}(r),\tag{3.31}$$

where $f_{\mathrm{closest}}(r)$ is the PDF of distance between the UE and its closest RRH already defined in equation (3.30), and $F_{\rho|r}$ is the CDF given by equation (3.21).

Next, by taking into consideration the fact that the BBU can transmit to the UE, which happens only when there is no RRH in cluster $\mathcal{G}$, we expand the definition of the nearest neighbour outage probability to the typical UE as

$$\mathbb{P}_{out}^{\mathrm{NNP}}(\xi) = \int_0^R f_{\mathrm{closest}}(r)F_{\rho_{k,0}|r_{k,0}}(\xi)\,\mathrm{d}r + \int_{\mathrm{R}}^{\infty} f_{\mathrm{closest}}(r)F_{\rho_0}(\xi)\mathrm{d}r,$$

where

$$F_{\rho_0}(\xi) = \left(1-\frac{1}{N_{\mathrm{t}}}\right)^{\mathcal{U}-1}\sum_{i\in\mathrm{L,N}}\frac{p_i\,\gamma\left(N_{\mathrm{t}}-K+1,\frac{\xi\,K\,R^{\alpha_i}}{B_{k,0}}\right)}{(N_{\mathrm{t}}-K)!}.\tag{3.32}$$

After substituting the expressions of $F_{\rho_0}(\xi)$, $F_{\rho_{k,0}|r_{k,0}}(\xi)$ and $f_{\mathrm{closest}}(r)$ into

equation (3.32), we obtain

$$\begin{aligned}\mathbb{P}_{out}^{\mathrm{NNP}}(\xi) \;=\; & \left(1-\frac{1}{N_{\mathrm{r}}}\right)^{\mathcal{U}-1} \underbrace{\int_0^R 2\pi\lambda_{\mathrm{RRH}}\, r e^{-\pi\lambda_{\mathrm{RRH}}\, r^2} \sum_{i\in\mathrm{L,N}} \frac{p_i\ \gamma\left(N_{\mathrm{r}}-K+1,\, \xi\, K\, r_{k,0}^{\alpha_i}\right)}{(N_{\mathrm{r}}-K)!}\,\mathrm{d}r}_{\mathrm{I}} \\ & + \left(1-\frac{1}{N_{\mathrm{t}}}\right)^{\mathcal{U}-1} \underbrace{\int_{\mathrm{R}}^{\infty} 2\pi\,\lambda_{\mathrm{RRH}}\, r e^{-\pi\,\lambda_{\mathrm{RRH}}\, r^2} \sum_{i\in\mathrm{L,N}} \frac{p_i\ \gamma\left(N_{\mathrm{t}}-K+1,\, \xi\, K\, R^{\alpha_i}\right)}{(N_{\mathrm{t}}-K)!}\,\mathrm{d}r}_{\mathrm{II}}\,.\end{aligned} \tag{3.33}$$

To integrate I in equation (3.33), we start with an LOS link, i.e.,

$$\begin{aligned}\mathrm{I_L} \;&=\; \left(1-\frac{1}{N_{\mathrm{r}}}\right)^{\mathcal{U}-1} \int_0^R 2\pi\lambda_{\mathrm{RRH}}\, r e^{-\pi\lambda_{\mathrm{RRH}}\, r^2} \sum_{i\in\mathrm{L,N}} \frac{p_{\mathrm{L}}\,\gamma\left(N_{\mathrm{r}}-K+1,\, \xi\, K\, r_{k,0}^{\alpha_{\mathrm{L}}}\right)}{(N_{\mathrm{r}}-K)!}\,\mathrm{d}r, \\ &\overset{(a)}{=}\; \left(1-\frac{1}{N_{\mathrm{r}}}\right)^{\mathcal{U}-1} \frac{p_{\mathrm{L}}}{(N_{\mathrm{r}}-K)!} \sum_{n=0}^{\infty} \frac{(-1)^n}{n!} \frac{(\xi\, K)^{(N_r-K+n+1)}}{(N_r-K+n+1)} \\ &\qquad\qquad \times \frac{\gamma\left(\frac{\alpha_{\mathrm{L}}(N_r-K+n+1)}{2}+1,\, \pi\,\lambda_{\mathrm{RRH}}\, R^2\right)}{(\pi\,\lambda_{\mathrm{RRH}})^{\frac{(\alpha_{\mathrm{L}}\,(N_r-K+n+1)}{2}}},\end{aligned} \tag{3.34}$$

where (a) follows from performing the integration after replacing the lower incomplete gamma function with its series equivalent given by

$$\gamma\left(N_{\mathrm{r}}-K+1,\, \xi\, K\, r_{k,0}^{\alpha_{\mathrm{L}}}\right) = \sum_{n=0}^{\infty} \frac{(-1)^n \left(\xi K r_{k,0}^{\alpha_{\mathrm{L}}}\right)^{N_{\mathrm{r}}+n-K+1}}{n!\,(Nr+n-K+1)}. \tag{3.35}$$

In a similar manner, $\mathrm{I_N}$ for the NLOS link can be derived. Subsequently, summing $\mathrm{I_N}$ and $\mathrm{I_L}$ yields the result of integrating I in equation (3.33) and is expressed as

$$\begin{aligned}\mathrm{I} = \left(1-\frac{1}{\mathrm{N_r}}\right)^{\mathcal{U}-1} & \frac{1}{(\mathrm{N_r}-\mathrm{K})!} \sum_{n=0}^{\infty} \frac{(-1)^{\mathrm{n}}}{\mathrm{n!}} \frac{(\xi\,\mathrm{K})^{(\mathrm{N_r}-\mathrm{K}+\mathrm{n}+1)}}{(\mathrm{N_r}-\mathrm{K}+\mathrm{n}+1)} \\ & \times \left[\sum_{i\in L,N} p_i\, \frac{\gamma\frac{(\alpha_i(N_r-K+n+1)}{2}+1, \pi\lambda_{\mathrm{RRH}}R^2)}{(\pi\lambda_{\mathrm{RRH}})^{\frac{(\alpha_i(N_r-K+n+1)}{2}}}\right].\end{aligned} \tag{3.36}$$

Next, integrating II, we have

$$\mathrm{II}=\left(1-\frac{1}{\mathrm{N_t}}\right)\sum_{\mathrm{i}\in\mathrm{L,N}}^{\mathcal{U}-1}\frac{\mathrm{p_i}\,\gamma\left(\mathrm{N_t}-\mathrm{K}+1,\,\xi\,\mathrm{K}\,\mathrm{R}^{\alpha_\mathrm{i}}\right)}{(\mathrm{N_t}-\mathrm{K})!}\;\mathrm{e}^{-\lambda_{\mathrm{RRH}}\,\pi\,\mathrm{R}^2}. \tag{3.37}$$

The proof of equation (3.29) is obtained from substituting equations (3.37) and (3.36) in equation (3.33).

3.4.2 BCP

In this scenario, the best channel in cluster $\mathcal{G}$ (either an RRH or the BBU) is selected to participate in transmission with the typical UE. This implies that outage occurs when this channel is in outage. Thus, the outage probability for the best channel is given as

$$\mathbb{P}_{out}^{\mathrm{BCP}}(\xi)=F_{\rho_0}(\xi)\,\mathbb{P}_{out}^{k}(\xi), \tag{3.38}$$

where $\mathbb{P}_{out}^{k}$ is the outage probability of the kth RRH given as

$$\mathbb{P}_{out}(\xi)=\left(1-\frac{1}{N_\mathrm{r}}\right)\sum_{i\in\mathrm{L,N}}^{\mathcal{U}-1}p_i\,\frac{2\,(\xi\,K)^{-\frac{2}{\alpha_i}}}{R^2}\left(1-\sum_{m=0}^{N_\mathrm{r}-K}\frac{\left[\Gamma\left(\frac{2}{\alpha_i}+m\right)-\Gamma\left(\frac{2}{\alpha_i}+m,KR^{\alpha_i}\xi\right)\right]}{m!}\right), \tag{3.39}$$

and $F_{\rho_0}(\xi)$ is the CDF of the SNR at the BBU given as

$$F_{\rho_0}(\xi)=\left(1-\frac{1}{N_\mathrm{t}}\right)^{\mathcal{U}-1}\sum_{i\in\mathrm{L,N}}\frac{p_i\,\gamma\left(N_\mathrm{t}-K+1,\frac{\xi\,K\,R^{\alpha_i}}{B_0}\right)}{(N_\mathrm{t}-K)!}. \tag{3.40}$$

3.5 Performance in an Interference-Limited Scenario

In the sequel, we present the corresponding outage analyses of the system whilst considering inter-cluster interference for both association scenarios.

3.5.1 BCP

In this scenario, the typical UE associates itself with the RRH that provides it with the best signal in the network. This can also be explained from the path-loss perspective where the best channel is the one that offers the least path-loss.

Proposition 3.3. *The outage probability of received SINR at the typical UE from the best RRH in the cluster is given as*

$$\mathbb{P}_{out}^{\mathrm{BCP}}(\xi) = \sum_{j\in \mathrm{L,N}} p_j \left(1-\frac{1}{N_{\mathrm{r}}}\right)^{\mathcal{U}-1} \sum_{k=0}^{\nu} \binom{\nu}{k} (-1)^k \tag{3.41}$$
$$\times \int_{y>0} e^{\frac{-A\,k\,\xi\,y\sigma_z^2}{G_0}} \prod_{j\in \mathrm{L,N}} \mathbb{E}_{\mathbf{I}^j_{\Phi'_{\mathrm{BS}}}} \left[\exp\left(\frac{-A\,k\,\xi\,y\mathbf{I}^j_{\Phi'_{\mathrm{BS}}}}{G_0}\right)\right] f_\zeta(y)\mathrm{d}y,$$

where $\nu = \mathrm{N}_r - \mathrm{K}$ *is a parameter from the tight upper bound of gamma distribution given as* $\mathbb{P}[|\eta_{\mathrm{m,n}}|^2 < \gamma < (1-e^{-A\gamma})^\nu]$ *with* $A = \nu(\nu!)^{\frac{-1}{\nu}}$ *and* $y \triangleq r_n^{\alpha_j}$*, and* f_ζ *is the distribution of the least path-loss, which is given by*

$$f_\xi(x) = \prod_{j\in\{L,N\}} \tfrac{2p_j}{\alpha_j} \pi \lambda_{\mathrm{RRH}} P_{\mathrm{RRH}}^{\frac{2}{\alpha_j}} x^{\frac{2}{\alpha_j}-1} e^{-\pi p_j \lambda_{\mathrm{RRH}} P_{\mathrm{RRH}}^{\frac{2}{\alpha_j}} x^{\frac{2}{\alpha_j}}}. \tag{3.42}$$

Proof. Let $y \triangleq r_n^{\alpha_j}$ represent the path loss. The outage probability conditioned on the least path-loss from the best RRH in the cluster to the typical UE averaged over the plane is defined as

$$\mathbb{P}_{\mathrm{out}|y}^{\mathrm{BCP}}(\xi) = \mathbb{E}\left[\mathbb{P}\left[\frac{G_0\,|\eta_{\mathrm{k},0}|^2\; y^{-1}\,\omega_p(N_{\mathrm{r}},\mathcal{U})}{\sigma_z^2 + \mathbf{I}_{\Phi'_{\mathrm{BS}}}} < \xi | y\right]\right]. \tag{3.43}$$

For the LOS link, $y = r^{\alpha_{\mathrm{L}}}$, the conditional outage probability is then given as

$$\begin{aligned}\mathbb{P}_{\mathrm{out}|y}^{\mathrm{L}}(\xi) &= \mathbb{E}\left[p_{\mathrm{L}}\,\mathbb{P}\left[\frac{G_0\,|\eta_{\mathrm{k},0}|^2\; y^{-1}\,\omega_p(N_{\mathrm{r}},\mathcal{U})}{\sigma_z^2 + \mathbf{I}_{\Phi'_{\mathrm{BS}}}} < \xi | y\right]\right], \qquad (3.44)\\ &= \int_{y>0} p_{\mathrm{L}}\,\mathbb{P}\left[\frac{G_0\,|\eta_{\mathrm{k},0}|^2\; y^{-1}\,\omega_p(N_{\mathrm{r}},\mathcal{U})}{\sigma_z^2 + \mathbf{I}_{\Phi'_{\mathrm{BS}}}} < \xi | y\right] f_\zeta(y)\,\mathrm{d}y.\end{aligned}$$

Given that the small-scale fading, $\eta_{k,0}$, is Rayleigh, $|\eta_{k,0}|^2$ follows chi-squared distribution with $2(N_{\mathrm{r}}-K)$ degrees of freedom, and employing the upper bound of gamma distribution with parameter ν such that $\mathbb{P}\left[|\eta_{k,0}|^2 < \gamma < (1-e^{-A\gamma})^\nu\right]$

with $A = \nu(\nu!)^{\frac{-1}{\nu}}$, the outage probability is expressed as

$$\begin{aligned}
&\mathbb{P}\left[\frac{G_0\,|\eta_{\mathrm{k},0}|^2\; y^{-1}\,\omega_p(N_{\mathrm{r}},\mathcal{U})}{\sigma_z^2+\mathbf{I}_{\Phi'_{\mathrm{BS}}}} < \xi|y\right] \\
&\quad= \mathbb{E}_{\mathbf{I}_{\Phi'_{\mathrm{BS}}}}\left[\mathbb{P}\left[|\eta_{\mathrm{k},0}|^2 < \frac{\xi\, y}{G_0\,\omega_p(N_{\mathrm{r}},\mathcal{U})}(\sigma_z^2+\mathbf{I}_{\Phi'_{\mathrm{BS}}})|y,\,\mathbf{I}_{\Phi'_{\mathrm{BS}}}\right]\right], \\
&\quad\overset{(a)}{=} \left(1-\frac{1}{N_{\mathrm{r}}}\right)^{\mathcal{U}-1}\mathbb{E}_{\mathbf{I}_{\Phi'_{\mathrm{BS}}}}\left[\left(1-e^{-A\,\frac{\xi\, y}{G_0}\,(\sigma_z^2+\mathbf{I}_{\Phi'_{\mathrm{BS}}})}\right)^{\nu}|y,\,\mathbf{I}_{\Phi'_{\mathrm{BS}}}\right], \\
&\quad\overset{(b)}{=} \left(1-\frac{1}{N_{\mathrm{r}}}\right)^{\mathcal{U}-1}\sum_{k=0}^{\nu}\binom{\nu}{k}(-1)^k e^{\frac{-A\,k\,\xi\,y\sigma_z^2}{G_0}}\mathbb{E}_{\mathbf{I}_{\Phi'_{\mathrm{BS}}}}\left[e^{-\frac{-A\,k\,\xi\,y\,\mathbf{I}_{\Phi'_{\mathrm{BS}}}}{G_0}}\right], \qquad (3.45)\\
&\quad\overset{(c)}{=} \left(1-\frac{1}{N_{\mathrm{r}}}\right)^{\mathcal{U}-1}\sum_{k=0}^{\nu}\binom{\nu}{k}(-1)^k e^{\frac{-A\,k\,\xi\,y\sigma_z^2}{G_0}}\prod_{j\in\mathrm{L,N}}\mathbb{E}_{\mathbf{I}^j_{\Phi'_{\mathrm{BS}}}}\left[\exp\left(\frac{-A\,k\,\xi\,y\,\mathbf{I}^j_{\Phi'_{\mathrm{BS}}}}{G_0}\right)\right],
\end{aligned}$$

where (a) follows from the precoding penalty and the tight gamma approximation previously defined, (b) follows from applying binomial expansion, and (c) follows from the fact that interference links can be LOS or NLOS such that $\mathbf{I}_{\Phi'_{\mathrm{BS}}} = \mathbf{I}^{\mathrm{L}}_{\Phi'_{\mathrm{BS}}} + \mathbf{I}^{\mathrm{N}}_{\Phi'_{\mathrm{BS}}}$. Substituting equation (3.45) into equation (3.44), we obtain the LOS outage probability as

$$\begin{aligned}
\mathbb{P}^{\mathrm{L}}_{\mathrm{out}}(\xi) &= \left(1-\frac{1}{N_{\mathrm{r}}}\right)^{\mathcal{U}-1} p_{\mathrm{L}}\int_{y>0}\sum_{k=0}^{\nu}\binom{\nu}{k}(-1)^k\, e^{\frac{-A\,k\,\xi\,y\sigma_z^2}{G_0}} \\
&\quad\times \prod_{j\in\mathrm{L,N}}\mathbb{E}_{\mathbf{I}^j_{\Phi'_{\mathrm{BS}}}}\left[\exp\left(\frac{-A\,k\,\xi\,y\,\mathbf{I}^j_{\Phi'_{\mathrm{BS}}}}{G_0}\right)\right] f_\zeta(y)\,\mathrm{d}y. \qquad (3.46)
\end{aligned}$$

Following similar steps, we can derive the NLOS outage probability $\mathbb{P}^{\mathrm{N}}_{\mathrm{out}}(\xi)$. Next, to obtain the least path-loss distribution f_ζ, let $\Theta = \{x = \frac{|\nu|_i^\alpha}{P_{\mathrm{RRH}}}, \nu \in \Phi_{\mathrm{RRH}}\}$ with intensity function λ where the subscript i of α_i denotes whether the link is LOS or NLOS. We find the intensity function λ and measure Λ of this process Θ that finds the RRH with the least path-loss by utilising the mapping theorem introduced in [92, Theorem 2.34] expressed as

$$\begin{aligned}
\Lambda &= \int_0^{P_{\mathrm{RRH}}^{\frac{1}{\alpha_i}}x^{\frac{1}{\alpha_i}}} 2\pi\lambda_{\mathrm{RRH}}\nu\mathrm{d}\nu, \\
&= \pi\lambda_{\mathrm{RRH}}P_{\mathrm{RRH}}^{\frac{2}{\alpha_i}}x^{\frac{2}{\alpha_i}}, \qquad (3.47)
\end{aligned}$$

and

$$\begin{aligned}\lambda &= \frac{\mathrm{d}}{\mathrm{d}x}(\Lambda) \\ &= \pi\lambda_{\mathrm{RRH}}\frac{2}{\alpha_i}P_{\mathrm{RRH}}^{\frac{2}{\alpha_i}}x^{\frac{2}{\alpha_i}-1}.\end{aligned} \tag{3.48}$$

Next, the distribution of the RRH with the least path-loss is obtain from the null probability of a PPP and given as

$$f_\zeta(x) = \lambda_{\mathrm{RRH}}(x)e^{-\Lambda_{\mathrm{RRH}}(x)}, \tag{3.49}$$

Substituting the values of λ and Λ from equations (3.48) and (3.47) respectively, we obtain the least path-loss distribution f_ζ given as

$$f_\zeta(x) = \prod_{j\in\{L,N\}} \tfrac{2p_j}{\alpha_j}\pi\lambda_{\mathrm{RRH}}P_{\mathrm{RRH}}^{\frac{2}{\alpha_j}}x^{\frac{2}{\alpha_j}-1}e^{-\pi p_j\lambda_{\mathrm{RRH}}P_{\mathrm{RRH}}^{\frac{2}{\alpha_j}}x^{\frac{2}{\alpha_j}}}. \tag{3.50}$$

To obtain the expectation of the LOS interfering link, we leverage results from [42, Lemma 6], focusing on the single-path case. Thus, the expectation is given as

$$\mathbb{E}_{\mathbf{I}^{\mathrm{L}}_{\Phi'_{\mathrm{BS}}}}\left[\exp\left(\frac{-A\,k\,\xi\,y\,\mathbf{I}^{\mathrm{L}}_{\Phi'_{\mathrm{BS}}}}{G_0}\right)\right] = \mathbb{E}_{\mathbf{I}^{\mathrm{L}}_{\Phi'_{\mathrm{BS}}}}\left[\prod_{b\in\Phi'_{\mathrm{BS}}}\exp\left(\frac{-A\,k\,\xi\,y\,G_{\mathrm{v}}\,x\Upsilon_v}{G_0\,r_v^{\alpha_{\mathrm{L}}}}\right)\right], \tag{3.51}$$

found by substituting $\mathbf{I}^{\mathrm{L}}_{\Phi'_{\mathrm{BS}}} = \sum_{b\in\Phi'_{\mathrm{BS}},v\neq 0} G_{\mathrm{v}}\left|\eta_{\mathrm{b,v}}\right|^2 \mathrm{r}_{\mathrm{v}}^{-\alpha_{\mathrm{L}}}$ and $x = \left|\eta_{\mathrm{b,v}}\right|^2$. Applying the probability-generating functional of PPP (PGFL) [92], we obtain

$$\mathbb{E}_{\mathbf{I}^{\mathrm{L}}_{\Phi'_{\mathrm{BS}}}}\left[\exp\left(\frac{-A\,k\,\xi\,y\,\mathbf{I}^{\mathrm{L}}_{\Phi'_{\mathrm{BS}}}}{G_0}\right)\right] = \exp\left(-2\,\pi\,\lambda_{\mathrm{RRH}}\int_r^\infty 1 - \frac{1}{\left(1+\frac{A\,k\,G_{\mathrm{v}}\,\xi\,y\tau^2_{\mathrm{RRH}}}{G_0\,r_v^{\alpha_{\mathrm{L}}}}\right)^{\nu}}\,p_{\mathrm{L}}\mathrm{d}r\right). \tag{3.52}$$

The expectation for an NLOS interfering link can be obtained similarly. Finally, the proof of outage probability from the best RRH in cluster $\mathcal{G}$ to the typical UE is obtained by the summation of both LOS and NLOS outage probabilities, respectively.

3.5.2 NNP

We consider the RRH closest to the typical UE, which experiences outage when the SINR is less than a predefined value.

Proposition 3.4. *The outage probability of the RRH closest to the typical UE considering the impact of inter-cluster interference is expressed as*

$$\begin{aligned}\mathbb{P}_{out}^{\mathrm{NNP}}(\xi) &= \sum_{j\in\mathrm{L,N}} p_j \left(1-\frac{1}{N_{\mathrm{r}}}\right)^{\mathcal{U}-1} \int_{r>0} \sum_{k=0}^{\nu} \binom{\nu}{k} (-1)^k \\ &\times e^{\frac{-A\,k\,\xi\,y\sigma_z^2}{G_0}} \prod_{j\in\mathrm{L,N}} \mathbb{E}_{\mathbf{I}^j_{\Phi'_{\mathrm{BS}}}} \left[\exp\left(\frac{-A\,k\,\xi\,y\mathbf{I}^j_{\Phi'_{\mathrm{BS}}}}{G_0}\right)\right] f_r(r)\,\mathrm{d}r,\end{aligned} \tag{3.53}$$

where y and ν are defined in equation (3.41) *and f_r is the PDF of the nearest distance to the typical UE given as*

$$f_r(r) = 2\,\pi\,\lambda_{\mathrm{RRH}}\,r\,e^{-\lambda_{\mathrm{RRH}}\,\pi\,r^2}. \tag{3.54}$$

Proof. The outage probability conditioned on the distance r from the nearest RRH in the cluster to the typical UE averaged over the plane is defined as

$$\mathbb{P}^{\mathrm{NNP}}_{\mathrm{out}|y}(\xi) = \mathbb{E}\left[\mathbb{P}\left[\frac{G_0\,|\eta_{\mathrm{k},0}|^2\;y^{-1}\,\omega_p(N_{\mathrm{r}},\mathcal{U})}{\sigma_z^2+\mathbf{I}_{\Phi'_{\mathrm{BS}}}} < \xi|r\right]\right]. \tag{3.55}$$

For the LOS link the conditional outage probability is then given as

$$\begin{aligned}\mathbb{P}^{\mathrm{L}}_{\mathrm{out}|r}(\xi) &= \mathbb{E}\left[p_{\mathrm{L}}\,\mathbb{P}\left[\frac{G_0\,|\eta_{\mathrm{k},0}|^2\;r^{-\alpha_{\mathrm{L}}}\,\omega_p(N_{\mathrm{r}},\mathcal{U})}{\sigma_z^2+\mathbf{I}_{\Phi'_{\mathrm{BS}}}} < \xi|r\right]\right], \\ &= \int_{r>0} p_{\mathrm{L}}\,\mathbb{P}\left[\frac{G_0\,|\eta_{\mathrm{k},0}|^2\;r^{-\alpha_{\mathrm{L}}}\,\omega_p(N_{\mathrm{r}},\mathcal{U})}{\sigma_z^2+\mathbf{I}_{\Phi'_{\mathrm{BS}}}} < \xi|r\right] f_r(r)\,\mathrm{d}r.\end{aligned} \tag{3.56}$$

Since the small-scale fading, $\eta_{k,0}$ follows Rayleigh fading, $|\eta_{k,0}|^2$ follows chi-squared distribution with $2(N_{\mathrm{r}}-K)$ degrees of freedom, and employing the upper bound of gamma distribution with parameter ν such that $\mathbb{P}\left[\,|\eta_{k,0}|^2 < \gamma < (1-e^{-A\gamma})^{\nu}\right]$

with $A = \nu(\nu!)^{\frac{-1}{\nu}}$, the outage probability is expressed as

$$\begin{aligned}
&\mathbb{P}\left[\frac{G_0\,|\eta_{\mathrm{k},0}|^2\; r^{-\alpha_{\mathrm{L}}}\,\omega_p(N_{\mathrm{r}},\mathcal{U})}{\sigma_z^2+\mathbf{I}_{\Phi'_{\mathrm{BS}}}} \;<\; \xi | r\right] \\
&\qquad = \mathbb{E}_{\mathbf{I}_{\Phi'_{\mathrm{BS}}}}\left[\mathbb{P}\left[|\eta_{\mathrm{k},0}|^2 \;<\; \frac{\xi\, r^{\alpha_{\mathrm{L}}}}{G_0\,\omega_p(N_{\mathrm{r}},\mathcal{U})}(\sigma_z^2+\mathbf{I}_{\Phi'_{\mathrm{BS}}}) | r,\, \mathbf{I}_{\Phi'_{\mathrm{BS}}}\right]\right], \\
&\qquad \overset{(a)}{=} \left(1-\frac{1}{N_{\mathrm{r}}}\right)^{\mathcal{U}-1}\mathbb{E}_{\mathbf{I}_{\Phi'_{\mathrm{BS}}}}\left[\left(1-e^{-A\,\frac{\xi}{G_0\,r^{\alpha_{\mathrm{L}}}}\,(\sigma_z^2+\mathbf{I}_{\Phi'_{\mathrm{BS}}})}\right)^{\nu} | r,\, \mathbf{I}_{\Phi'_{\mathrm{BS}}}\right], \\
&\qquad \overset{(b)}{=} \left(1-\frac{1}{N_{\mathrm{r}}}\right)^{\mathcal{U}-1}\sum_{k=0}^{\nu}\binom{\nu}{k}(-1)^k e^{\frac{-A\,k\,\xi\,r^{-\alpha_{\mathrm{L}}}\sigma_z^2}{G_0}}\mathbb{E}_{\mathbf{I}_{\Phi'_{\mathrm{BS}}}}\left[e^{-\frac{-A\,k\,\xi\,r^{-\alpha_{\mathrm{L}}}\,\mathbf{I}_{\Phi'_{\mathrm{BS}}}}{G_0}}\right], \qquad (3.57) \\
&\qquad \overset{(c)}{=} \left(1-\frac{1}{N_{\mathrm{r}}}\right)^{\mathcal{U}-1}\sum_{k=0}^{\nu}\binom{\nu}{k}(-1)^k e^{\frac{-A\,k\,\xi\sigma_z^2}{G_0\,r^{\alpha_{\mathrm{L}}}}}\prod_{j\in\mathrm{L},\mathrm{N}}\mathbb{E}_{\mathbf{I}^j_{\Phi'_{\mathrm{BS}}}}\left[\exp\left(\frac{-A\,k\,\xi\,\mathbf{I}^j_{\Phi'_{\mathrm{BS}}}}{G_0\;r^{\alpha_{\mathrm{L}}}}\right)\right],
\end{aligned}$$

where (a) follows from the precoding penalty and the tight gamma approximation previously defined, (b) follows from applying binomial expansion, and (c) follows from the fact that interference links can be LOS or NLOS such that $\mathbf{I}_{\Phi'_{\mathrm{BS}}} = \mathbf{I}^{\mathrm{L}}_{\Phi'_{\mathrm{BS}}} + \mathbf{I}^{\mathrm{N}}_{\Phi'_{\mathrm{BS}}}$. Next, substituting equations (3.57) and (3.54) into equation (3.56), we obtain the LOS outage probability as

$$\begin{aligned}
\mathbb{P}^{\mathrm{L}}_{\mathrm{out}}(\xi) \;&=\; \left(1-\frac{1}{N_{\mathrm{r}}}\right)^{\mathcal{U}-1} 2\pi\lambda_{\mathrm{RRH}}p_{\mathrm{L}}\int_{r>0}\sum_{k=0}^{\nu}\binom{\nu}{k}(-1)^k\, e^{\frac{-A\,k\,\xi\sigma_z^2}{G_0\,r^{\alpha_{\mathrm{L}}}}} \\
&\quad\times\; \prod_{j\in\mathrm{L},\mathrm{N}}\mathbb{E}_{\mathbf{I}^j_{\Phi'_{\mathrm{BS}}}}\left[\exp\left(\frac{-A\,k\,\xi\mathbf{I}^j_{\Phi'_{\mathrm{BS}}}}{G_0\,r^{\alpha_{\mathrm{L}}}}\right)\right] re^{-\lambda_{\mathrm{RRH}}\pi r^2}\mathrm{d}r. \qquad (3.58)
\end{aligned}$$

By following similar steps, we can derive the NLOS outage probability as

$$\begin{aligned}
\mathbb{P}^{\mathrm{N}}_{\mathrm{out}}(\xi) \;&=\; \left(1-\frac{1}{N_{\mathrm{r}}}\right)^{\mathcal{U}-1} 2\pi\lambda_{\mathrm{RRH}}p_{\mathrm{N}}\int_{r>0}\sum_{k=0}^{\nu}\binom{\nu}{k}(-1)^k\, e^{\frac{-A\,k\,\xi\sigma_z^2}{G_0\,r^{\alpha_{\mathrm{N}}}}} \\
&\quad\times\; \prod_{j\in\mathrm{L},\mathrm{N}}\mathbb{E}_{\mathbf{I}^j_{\Phi'_{\mathrm{BS}}}}\left[\exp\left(\frac{-A\,k\,\xi\mathbf{I}^j_{\Phi'_{\mathrm{BS}}}}{G_0\,r^{\alpha_{\mathrm{L}}}}\right)\right] re^{-\lambda_{\mathrm{RRH}}\pi r^2}\mathrm{d}r. \qquad (3.59)
\end{aligned}$$

Finally, the proof of outage probability from the nearest RRH to the typical UE can be obtained by the summation of both LOS and NLOS outage probabilities, respectively.

Remark 3.2. *Note that the throughput for both UE association scenarios in noise-limited and interference-limited cases can be obtained by the direct appli-*

cation of the respective outage probabilities derived in sections 3.4 and 3.5 to equation (3.28).

3.6 Network Latency

In this section, we investigate the average latency as a measure of the QoS of the network.

3.6.1 Average rate

In order to compute the average latency, we need to characterise the average rate. Therefore in this subsection we adopt the framework developed by the work in [93] to evaluate average rate in terms of moment-generating functions (MGFs). Thus, we present the average rate between the typical UE and its associated RRH in the next proposition[12] [13]

Proposition 3.5. *The average rate between a typical UE and its associated RRH is given as*

$$\bar{\mathcal{R}} = \left(1 - \frac{1}{N_{\mathrm{r}}}\right)^{\mathcal{U}-1} \int_0^{\infty} (1 - \mathcal{L}_S(t))\ \mathcal{L}_{\mathbf{I}_{\Phi'_{\mathrm{BS}}}}(t) \frac{e^{-t}}{t}\, \mathrm{d}t, \tag{3.60}$$

where S denotes the SNR and $\mathbf{I}_{\Phi'_{\mathrm{BS}}}$ represents the interference,

$$\mathcal{L}_S(t) = \sum_{k=0}^{\nu} \binom{\nu}{k} (-1)^{k+1} \left(1 + \frac{t\, G_0}{A\, k\, r_{k,0}^{\alpha_{\mathrm{L}}} \sigma_z^2}\right)^{-1} p_{\mathrm{L}} + \sum_{k=0}^{\nu} \binom{\nu}{k} (-1)^{k+1} \left(1 + \frac{t\, G_0}{A\, k\, r_{k,0}^{\alpha_{\mathrm{N}}} \sigma_z^2}\right)^{-1} p_{\mathrm{N}}, \tag{3.61}$$

and

$$\mathcal{L}_{\mathbf{I}_{\Phi'_{\mathrm{BS}}}}(t) = \prod_{q \in \mathrm{L,N}} p_q \exp\left(- 2\pi \lambda_{\mathrm{RRH}} \int_r^{\infty} 1 - \left(1 + \frac{A\, k\, t\, G_{\mathrm{v}}\, r_{k,0}^{\alpha_q}}{G_0\, r_v^{\alpha_q}}\right)^{-\nu} \mathrm{d}r \right).$$

Proof. From the generalised expression for average rate in terms of MGFs

[12] Note that this proposition also holds for the noise-limited case when the interference is equal to zero.

[13] It is important to note that although propositions 3.3-3.5 are not in closed form, they give an approximation for their respective outage probabilities and average rate performances, providing useful analytical insights into the performance of the mmWave CRAN system.

[93],

$$\bar{\mathcal{R}} = \int_0^{\infty} (1 - \mathcal{L}_S(x)) \mathcal{L}_{\mathbf{I}_{\Phi'_{\mathrm{BS}}}}(x) \frac{e^{-x}}{x} \, \mathrm{d}x, \tag{3.62}$$

where

$$S = \frac{G_0 \left|\eta_{\mathrm{k},0}\right|^2 \mathrm{r}_{k,0}^{-\alpha_j}}{\sigma_z^2}, \tag{3.63}$$

and

$$\mathbf{I}_{\Phi'_{\mathrm{BS}}} = \sum_{b \in \Phi'_{\mathrm{BS}}, v \neq 0} G_{\mathrm{v}} \left|\eta_{\mathrm{b,v}}\right|^2 \mathrm{r}_{\mathrm{v}}^{-\alpha} \sum_{v \in \mathcal{U}_v} ||\mathbf{a}_{\mathrm{UE}}(\theta_{v,y,x}) \mathbf{a}_{\mathrm{RRH}}^{\mathrm{H}}(\phi_{v,y,x})||^2. \tag{3.64}$$

The MGF of S is obtained after following similar steps from the proof of proposition 3.3 and is expressed as

$$\mathcal{L}_S(t) = \sum_{k=0}^{\nu} \binom{\nu}{k} (-1)^{k+1} \left(1 + \frac{t\, G_0}{A\, k\, r_{k,0}^{\alpha_{\mathrm{L}}} \sigma_z^2}\right)^{-1} p_{\mathrm{L}} + \sum_{k=0}^{\nu} \binom{\nu}{k} (-1)^{k+1} \left(1 + \frac{t\, G_0}{A\, k\, r_{k,0}^{\alpha_{\mathrm{N}}} \sigma_z^2}\right)^{-1} p_{\mathrm{N}}.$$

In like manner, the MGF of the interference-to-noise ratio, $\mathcal{L}_{\mathbf{I}_{\Phi'_{\mathrm{BS}}}}(t)$ in equation (3.62) is obtained using steps similar to equations (3.51)–(3.52). This concludes the proof.

Remark 3.3. *The analysis of this chapter focuses on the access links between RRHs and UEs. However, it is worthy of mention that the fronthaul rate is constant and dependent on digitised I and Q samples and the number of antennas as described in section 4.4.1 of [94], and is given as*

$$\mathcal{R}_{\mathrm{FH}} = 2\gamma f_{\mathrm{s}} N_{\mathrm{A}} N_q, \tag{3.65}$$

where the factor 2 accounts for I and Q phases of the signal, γ represents the overhead introduced by FEC and control signals, N_{A} is the number of antennas, and f_{s} and N_{q} are the sampling frequency and resolution of the quantiser, respectively.

3.6.2 Traffic latency

In this subsection, we utilise the queuing theory model described in 3.2.4 to analyse the delay in data delivery from BBU to UE_0.

3.6.2.1 Access link latency

In the network, where an RRH is connected to different UEs, if we assume that each UE connected to an RRH is served in a round-robin manner, and denote the coverage area of all RRHs as $\mathcal{B}$, then the average traffic load density at UE_0 from RRH_k is defined as

$$\delta(0) = \frac{\lambda_0 \, \kappa_0}{\bar{\mathcal{R}}_0}, \tag{3.66}$$

where 0 denotes the origin of the cluster where UE_0 is located, $\bar{\mathcal{R}}_0$ is the access link rate defined in equation (3.60), and κ_0 is the traffic load of UE_0 associated with RRH_k given by

$$\kappa_0 = \frac{\lambda_0}{\mu_0}, \tag{3.67}$$

where λ_0 is the arrival rate per unit area and μ_0 is the average number of requested volumes (average traffic load).

Therefore, the total traffic load in RRH_k is expressed as

$$D_k = \int_{\mathcal{B}} \delta(x) \, \mathrm{d}x, \tag{3.68}$$

where x denotes the location of the UE connected to RRH_k and $\delta(x)$ denotes the average traffic load density of the UE at location x. Next, computing the required service time to satisfy the demands of UE_0, we have

$$\nu_0 = \frac{\kappa_0}{\mathcal{R}_0}. \tag{3.69}$$

Given that traffic delivery in the access link emulates the M/G/1 queueing model, the average traffic delivery time for UE_0 in RRH_k is given as [85]

$$T_k = \frac{\kappa_0}{\mathcal{R}_0(1 - D_k)}. \tag{3.70}$$

At RRH_k, the average waiting time for the traffic load of UE_0 is obtained by subtracting the required service time from the average traffic delivery time and is expressed as

$$\chi_k = T_k - \nu_0 = \frac{D_k \, \kappa_0}{\mathcal{R}_0 \, (1 - D_k)}; \tag{3.71}$$

the latency in the access link of RRH_k and UE_0 is then calculated as the ratio of

waiting time and service time. Accordingly,

$$\Psi_k = \frac{\chi_k}{\nu_0} = \frac{D_k}{1 - D_k}. \tag{3.72}$$

3.6.2.2 Fronthaul-link latency

In the fronthaul link between the BBU and RRH_k, the required time to satisfy the traffic demand of UE_0 is dependent on the fronthaul rate, $\mathcal{R}_{\text{FH}}$. Thus the required time is expressed as

$$\hat{\nu}_0 = \frac{\kappa_0}{\mathcal{R}_{\text{FH}}}, \tag{3.73}$$

where $\mathcal{R}_{\text{FH}}$ is defined in equation (3.65). From the M/M/1 queueing model in [85], the average wait time for UE_0's traffic load in RRH_k's fronthaul is defined as

$$\hat{\chi}_k = \frac{\hat{D}_k \, \kappa_0}{\mathcal{R}_{\text{FH}} \, (1 - \hat{D}_k)}, \tag{3.74}$$

where $\hat{D}_k$ is the total load in the fronthaul of RRH_k. In this case, the latency ratio to measure how much time UE_0 waits per unit service time in the fronthaul of RRH_k is given as

$$\hat{\Psi}_k = \frac{\hat{\chi}_k}{\hat{\nu}_0} = \frac{\hat{D}_k}{1 - \hat{D}_k}. \tag{3.75}$$

Consequently, the traffic-delivery latency from BBU to UE_0 via RRH_k is given as

$$\Delta_k = \hat{\Psi}_k + \Psi_k. \tag{3.76}$$

Remark 3.4. *Note that the latency ratio of the fronthaul link in equation* (3.75) *is dependent only on the traffic load at the fronthaul of RRH_k, implying that all UEs associated with RRH_k have the same latency ratio. Therefore, a smaller Δ_k would suggest that RRH_k introduces a low latency to its associated UEs.*

3.7 Numerical Results

This section provides numerical results to validate the system model and analytical derivations from the aforementioned propositions. In Figs 3.3–3.7, we show how the performance of the outage probability varies to system factors such as RRH density, cluster radius, path-loss exponents and blockage density, while Figs 3.9 and 3.10 illustrate the throughput and average latency performances.

Table 3.1: Multi-user mmWave CRAN simulation parameters.

Notation	Parameter	Value
R	Cluster radius	250 meters
N_{UE}	Number of UE antennas	2×2, 4×4 UPA
B	Bandwidth	2GHz
N_{r}	Number of RRH antennas	4×4,8×8 UPA
N_{T}	Number of BBU antennas	4×4, 8×8 UPA
K	Number of UEs	4
p_{L}, R	Blockage model	0.1, 250 meters
P_T	Transmit power	30 dBm
λ_{RRH}	Density of RRH nodes	$6/1000\pi$
α	Path-loss exponent	L-2, N-3.5
N_0	Noise power	-100 dBm

The parameters used for all the simulations, unless otherwise stated, are tabulated in Table 3.1 with their corresponding values taken from literature mentioned in the references. We evaluate the RRH intensity as $\lambda_{\mathrm{RRH}} = \dfrac{\Lambda}{\pi R^2}$, where Λ represents the average number of RRHs in a cluster of size πR^2 with transmit power of 30 dBm and noise power of -100 dBm. A cluster radius of 250 m is considered.

Starting with the BCP scenario of the noise-limited case from equation (3.38), Fig. 3.3 presents a plot of outage probability against SNR threshold, and we observe that with an increase in the required threshold, the probability of outage also increases. However, as the average number of RRHs in the cluster is increased, outage probability is reduced. We also show the validity of the approximations used in the analysis of the outage probability by the small difference between the simulation and analytical results. An interesting observation is the impact of inter-cluster interference on the outage probability. This will subsequently be explained from Fig. 3.6.

Next, we consider the effect of changing the cluster radius on outage probability in Fig. 3.4. Here, we observe that increasing the radius does not increase the outage probability for the best channel scenario. We may infer by this result that there is a trade-off between cluster radius and outage performance in mmWave CRAN, as smaller cluster radii lead to better performance. This can be used in network planning by operators, especially in urban areas where many small

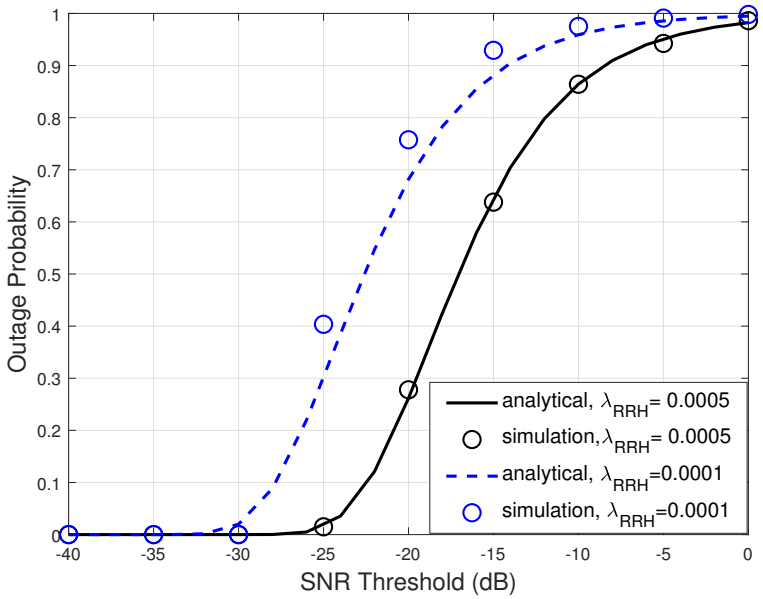


Figure 3.3: Comparison between analytical and simulated BCP outage probability in a noise-limited scenario under varying RRH densities.

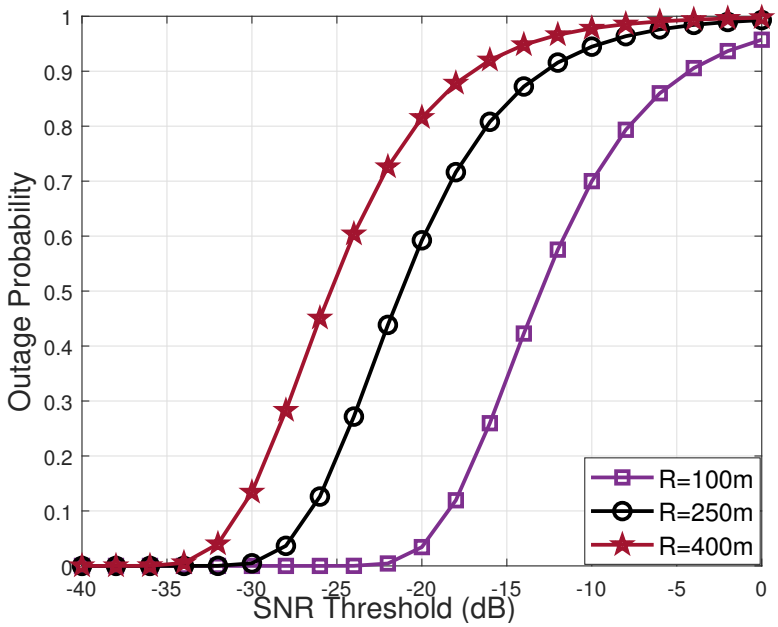


Figure 3.4: BCP outage probability against SNR in a noise-limited scenario for varying cluster radii.

clusters can be formed.

Fig. 3.5 shows the comparison of outage probability in both noise-limited and

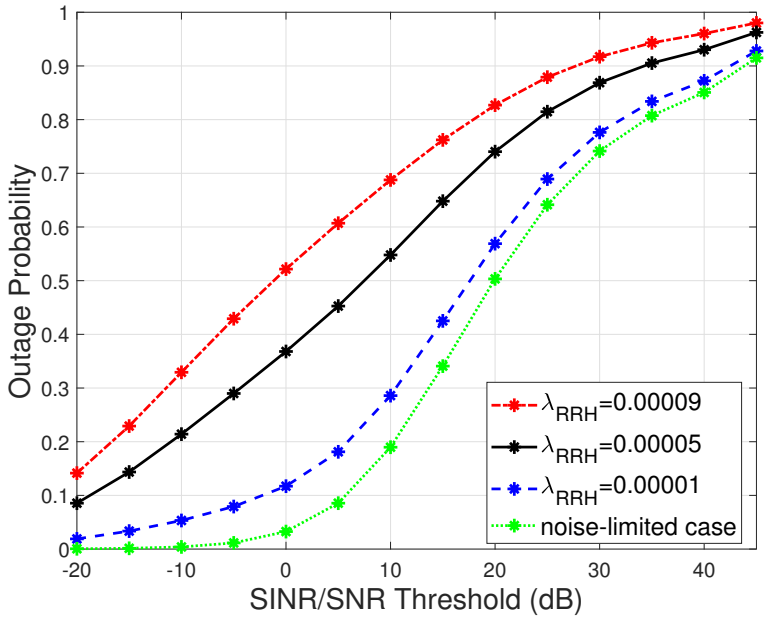


Figure 3.5: Comparison of NNP outage probability for noise-limited and interference-limited scenarios under different RRH densities.

interference-limited scenarios for NNP(i.e., comparing propositions 3.2 and 3.4.). It is evident from the figure that for a given moderate number of blockages, the outage probability is much less in the noise-limited regime when compared to the interference-limited regime. It can also be observed that as the number of interferers is decreased (by the reduction in RRH node density), the performance in the interference-limited scenario tends to that in the noise-limited regime. This outcome indicates that in systems employing mmWave links, successful transmission is largely dependent on blockage and nodal densities.

Fig. 3.6 shows that an increase in the average number of RRHs does not lead to a decrease in outage probability from proposition 3.3. Although this result appears to be counter-intuitive, it can be explained by the fact that increasing the number of RRHs also increases the probability of interfering RRHs. We establish by the small difference in outage probabilities between RRH intensities in Fig. 3.6, a trade-off between outage performance and interference. It is also evident from Fig. 3.6 that increasing the array size improves the performance, leading to smaller outage probabilities.

In Fig. 3.7, we plot the outage probability as a function of SINR threshold. In measuring the effect of path loss on the transmission from the best channel to a typical UE, we observe that outage probability decreases with reducing α. We

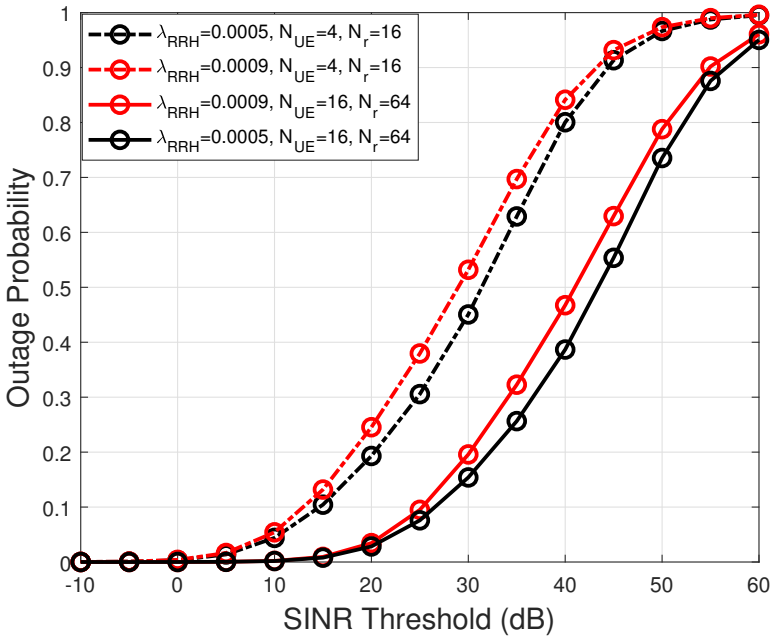


Figure 3.6: Comparison of BCP outage probability in an interference-limited scenario under different RRH densities and number of antennas.

may thus deduce that increasing the value of the path-loss exponent will degrade communication.

Next, Fig. 3.8 illustrates the outage probability of the BCP and NNP scenarios against SINR threshold comparing propositions 3.3 and 3.4. It is clear that BCP always outperforms NNP. For example, to achieve an SINR threshold of 20 dB when RRH density is 0.0001, the outage probabilities for BCP and NNP are 0.25 and 0.8, respectively. Although NNP is considered in some CRAN applications for the reduction of overhead in the selection process, it is worthwhile to note that in multi-user mmWave CRAN systems we opt for the best channel selection process for better transmission due to the big difference between NNP and BCP performances.

Fig. 3.9 illustrates the throughput performance of BCP and NNP schemes against SINR threshold by applying equation (3.28). In Fig. 3.9, we show the comparison of throughput performance for both BCP and NNP scenarios. Following from the result that increasing the number of RRHs can cause a degradation in communication between either the best or nearest RRH and the typical UE, the throughput performance for an RRH intensity of 0.0001 is better than that for an intensity of 0.0005 in both cases. Furthermore, Fig. 3.9 shows that BCP

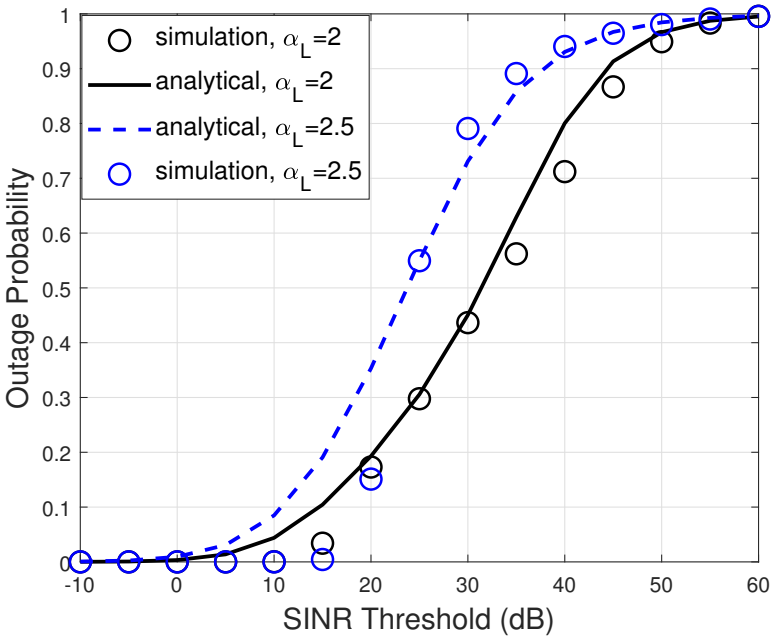


Figure 3.7: Comparison of BCP outage probability in an interference-limited scenario with varying path-loss exponents.

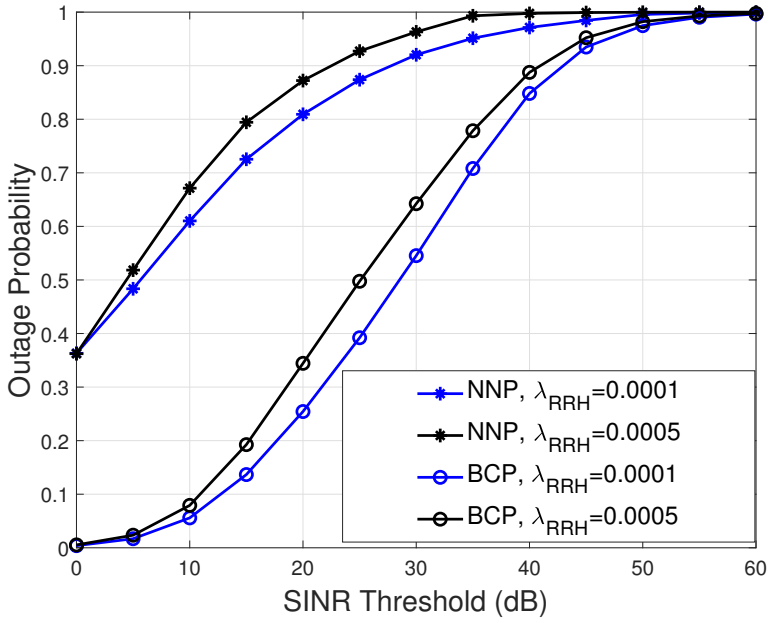


Figure 3.8: Outage probability comparison of BCP and NNP in an interference-limited scenario.

significantly outperforms NNP. To achieve an SINR of 20 dB, the BCP throughput for $\lambda_{\mathrm{RRH}} = 0.0001$ is 5 bits/Hz while that of NNP is 1.3 bits/Hz; this confirms

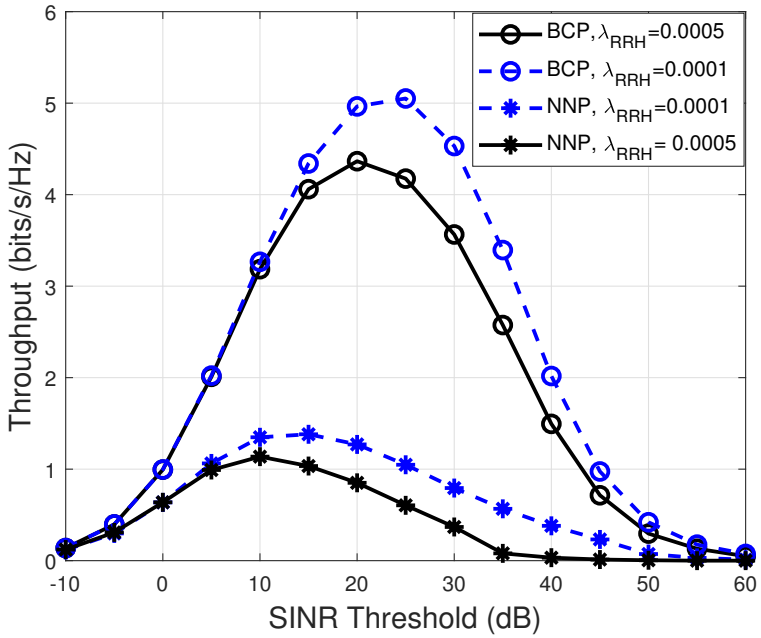


Figure 3.9: Throughput comparison of BCP and NNP in an interference-limited scenario.

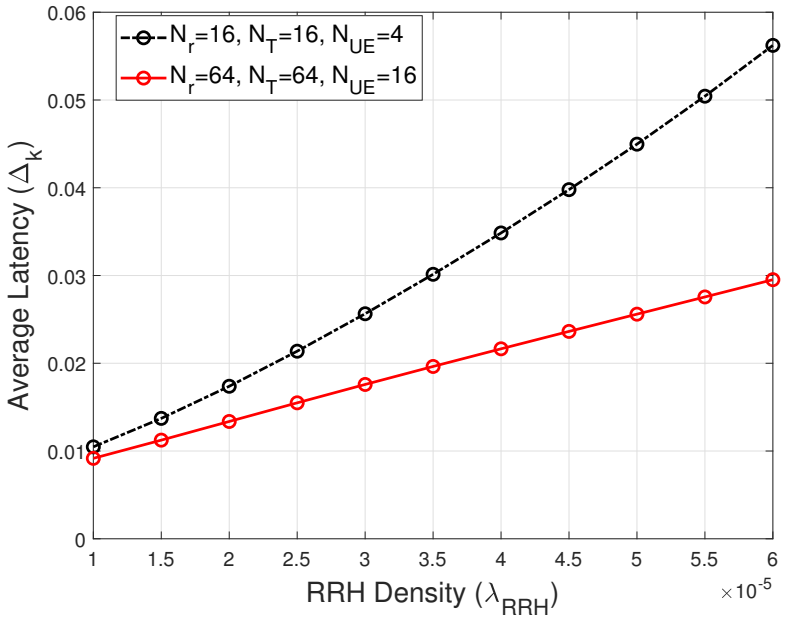


Figure 3.10: Comparison of average latency performance against RRH density with varying fronthaul rates.

the results of Fig. 3.8. Note that there exists an optimal value of throughput as shown by the shape of the curves in Fig. 3.9, with the implication that increas-

ing SINR does not indefinitely lead to better performance. Determination of the optimal point can be explored in future works.

Having seen the effect on increasing the number of RRHs on outage probability, we now determine the impact of varying RRH density on the average data-delivery latency in the network. Hence, Fig. 3.10 is plotted for different values of λ_{RRH} obtained from equation (3.76). We observe that increasing the RRH density leads to a rise in average latency. This outcome, although unexpected, is not so unusual given that an increased number of RRHs implies a larger traffic load at the fronthaul queue, leading to more latency. In addition, we also observe that increasing the antenna gain, which in turn increases the fronthaul and access link rates, leads to reducing the latency. It is important to note that at low RRH density, the latency for both antenna gains are almost equal due to the fact that when the traffic load is small, there is no significant gain in increasing the fronthaul or access link rate. In other words, the impact of fronthaul and access link rates are negligible for small traffic loads.

3.8 Conclusion

In this chapter, we studied the performance of a system that couples mmWave and CRAN (mmWave CRANs) for future-generation communication. We considered the downlink scenario of multiple distributed RRHs and a BBU in a multi-user system employing mmWave technologies such as hybrid beamforming in both fronthaul and access links. Analytical expressions for outage probability and throughput were derived for two UE association scenarios, namely, BCP and NNP. These analytical expressions provide a means of measurement for the performance of a mmWave CRAN system. We considered two cases to model the practical deployment of mmWave CRANs networks; namely, (a) noise-limited and (b) interference-limited. In addition, we analysed the impact of the fronthaul rate on the average data-delivery latency in the system. These analytical expressions were validated by simulations in the section of numeral results. Specifically, our results showed that there exists a trade-off between RRH deployment (measured by the density) and inter-cluster interference. Additionally, we showed that deploying larger antenna arrays can compensate for the degradation of communication in terms of delivery latencies, throughput and outage probabilities with higher RRH deployment. This can be exploited by engineers in practice for the

maintenance of high performance. We also showed that in highly dense urban environments, there can be a positive impact of increased blockages and path-loss exponents on outage probability and throughput in a highly dense deployment . For the UE association scenarios, our results clearly showed that BCP scenarios grossly outperform the NNP scenarios. Despite the attractiveness of NNP in existing literature, based on its reduction in overheads compared to other scenarios, the BCP is shown to be more viable for multi-user mmWave CRAN systems.

Chapter 4

Energy-efficient transceiver design in mmWave systems

4.1 Introduction

In Chapter 3, we presented a functional network architecture for the multi-user mmWave cellular communication with hybrid beamforming and analysed its performance. In this chapter, we propose energy-efficient hybrid beamformers which enable us to address the energy-efficiency concerns in the practical implementation of mmWave systems, both in the combination of mmWave and massive MIMO (outlined earlier in section 2.1.1), and in the mmWave-connected devices.

Recent years have seen a paradigm shift in the definition of connected devices, and as a result one of the requirements of future wireless networks is that they must be robust to exponential increases in the number of connected devices [95]. IoT is an emerging, paradigm-shifting technology that enables the connectivity of physical things in a network, so that communication is possible between devices such as cars, bicycles, speakers, headphones, fitness trackers, etc. With applications in many different sectors, such as healthcare, transportation, industry and home automation, IoT will form an integral part of future wireless networks [27, 96]. However, providing such interconnectivity will require a network with enormous amounts of bandwidth to support the exponentially increasing number of connected devices.

Work in this chapter is under minor revision for publication in IEEE Transactions on Green Communications and Networking, January 2019.

Appropriately, the large amounts of unused/moderately used bandwidths in the mmWave spectrum can be used to achieve IoT communication. In other words, the mmWave spectrum has the potential to provide a significant increase in capacity and support millions of connected devices with high reliability and reduced latency [3]. Based on the facts that mmWave systems require large antenna arrays to achieve high quality communication [13, 21], and that IoT devices are mostly energy-constrained, it becomes infeasible to employ digital processing (requiring an RF chain per antenna) for IoT communications at such frequencies. Consequently, hybrid precoding and combining (a general overview of which was provided earlier in section 2.3.2) can be utilised in IoT communications.

Within this context, in this chapter we consider a hybrid transceiver design problem with both perfect and imperfect CSI considerations for mmWave communications involving IoT devices. In general, the efficiency of a communication system is measured in terms of spectral efficiency (SE). This metric evaluates how the frequency spectrum is being utilised and, as such, most prior work on designing the hybrid precoders in mmWave communication [21, 28, 42, 80, 97, 98, 82] focus on SE maximisation using iterative maximisation or minimisation algorithms. However, if we consider the fact that IoT devices are more constrained in energy than capacity, the above design methodologies can be considered limited since they provide no insight on the efficiency of energy consumption in the network.[1] Nevertheless, some recent works on hybrid precoding do consider energy efficiency (EE) as a design criteria [43, 82, 99]. However, they do so by implementing a partially-connected structure, using a reduced number of phase shifters, and providing algorithms that reduce the rate loss characteristic of the structure. Our work differs from theirs in that we investigate the non-trivial fully-connected hybrid transceiver design for EE maximisation in mmWave IoT networks. It is important to note that although the partially-connected hybrid tranceiver design is less complicated than the fully-connected structure, it incurs a significant capacity loss of $N \log_2 N$ bits/s/Hz [43]. Therefore, designing an energy-efficient fully-connected hybrid transceiver is of interest and is the focus of this work.

Furthermore, a commonality among the aforementioned studies on hybrid transceiver design is the assumption of ideal hardware. In practice however, low-power circuits, which are more sensitive to hardware impairments, are de-

[1] The terms "network" and "system" are used interchangeably throughout this chapter; both represent the same thing.

ployed at the transceivers of mmWave systems for cost-effective implementation of large-scale antenna arrays. In such cases, signal processing in the digital domain fails to eliminate the impairments that arise from quantization errors, in-phase/quadrature (I/Q) imbalance, non-linearities in the amplifiers and mutual coupling between antenna ports [100, 101]. Therefore, robust mmWave transceiver designs which take into consideration the hardware impairments impacting the system's performance are of considerable interest.

To this end, in this chapter, we focus on the robust hybrid transceiver design for EE maximisation in mmWave interference channels involving IoT devices. We propose a two-stage algorithm to obtain the near-optimal hybrid transceiver design solutions. In the first stage, we first transform the hybrid transceiver problem into its univariate equivalent and propose a two-layer algorithm for solving the univariate problem. For the inner layer, we transform the fractional objective function into a subtractive equivalent, and exploit the relationship between WSR and WMMSE problems to solve efficiently for the optimal precoders and combiners, while in the outer layer, we apply a one-dimensional bisection search to determine the optimal EE parameter. On the other hand, in the second stage, we separate the design of the hybrid precoders and combiners into two sub-problems and adopt an orthogonal matching pursuit approach to solve each optimisation sub-problem. This technique is guaranteed to converge and has been proven to work for WSR maximisation in mmWave systems [28, 42, 97]. Furthermore, the hybrid transceiver design is improved to account for imperfect CSI based on the stochastic error modelling. The main distinctions of this chapter are summarised in terms of design guidelines, as follows.

- The power-constrained nature of ubiquitous smart devices make EE a crucial design criteria in IoT communication. In a mmWave interference channel involving IoT devices, the problem is further elevated by the inter-device interference. Thus, we consider the non-trivial EE maximisation transceiver design problem.

- In hybrid transceivers, which combine both analog and digital processing, additional constraints involving multiple variables as well as the non-convex constraints on the analog precoders must be considered.

- In a mmWave IoT network, the quality of communication, affected by the inevitable transceiver impairments, strongly relates to the EE per-

formance/design of the system. In this regard, we consider an accurate transceiver model that accounts for the transmitter and receiver distortions.

- In IoT networks, the achievable EE may be affected by the inevitable CSI errors at the transmitters. Hence, it is important to look at the resilience of the transceiver designs for error-prone CSI estimates at the transmitter.

The rest of this chapter is organised as follows. Section 4.2 presents some preliminaries including the system model, power consumption model and EE maximisation problem formulation. Section 4.3 provides the equivalent EE univariate problems and solutions. In Section 4.4, we tackle the hybrid precoder and combiner sub-problems and provide insights on the complexity of the proposed algorithms, while in section 4.5, we provide the robust transceiver designs under imperfect CSI considerations. Numerical results are presented in section 4.6. Finally, section 4.7 presents the conclusion of the chapter.

4.2 Preliminaries

4.2.1 System model

We consider K–user interference channels consisting of $2K$ mmWave IoT devices, K of which are transmitting while the other K represents the receiving devices. For simplicity, we assume that each transmitting device is equipped with a hybrid processor and is paired with a single receiving device in a one-to-one mapping as illustrated in Fig. 4.1. Specifically, each transmitter with N_{T} antennas and $N_{\mathrm{T}}^{\mathrm{RF}}$ RF chains communicates with its corresponding receiver with N_{R} antennas and $N_{\mathrm{R}}^{\mathrm{RF}}$ RF chains by sending N_{s} independent data streams. The conditions to enable multi stream communication and satisfy the degrees of freedom of interference networks are described by [102]

$$N_{\mathrm{s}} = \min(N_{\mathrm{T}}^{\mathrm{RF}}, N_{\mathrm{R}}^{\mathrm{RF}}), \tag{4.1}$$

$$N_{\mathrm{s}} \leq N_{\mathrm{T}}^{\mathrm{RF}} \leq N_{\mathrm{T}}, \tag{4.2}$$

$$N_{\mathrm{s}} \leq N_{\mathrm{R}}^{\mathrm{RF}} \leq N_{\mathrm{R}}. \tag{4.3}$$

The hybrid processor at the transmitting device j consists of an $N_{\mathrm{T}}^{\mathrm{RF}} \times N_{\mathrm{s}}$ baseband precoder $\mathbf{V}_{j,\mathrm{BB}}$, and an $N_{\mathrm{T}} \times N_{\mathrm{T}}^{\mathrm{RF}}$ RF precoder $\mathbf{V}_{j,\mathrm{RF}}$, as illustrated in Fig. 4.2. Hence, the transmitted signal from device j is given by

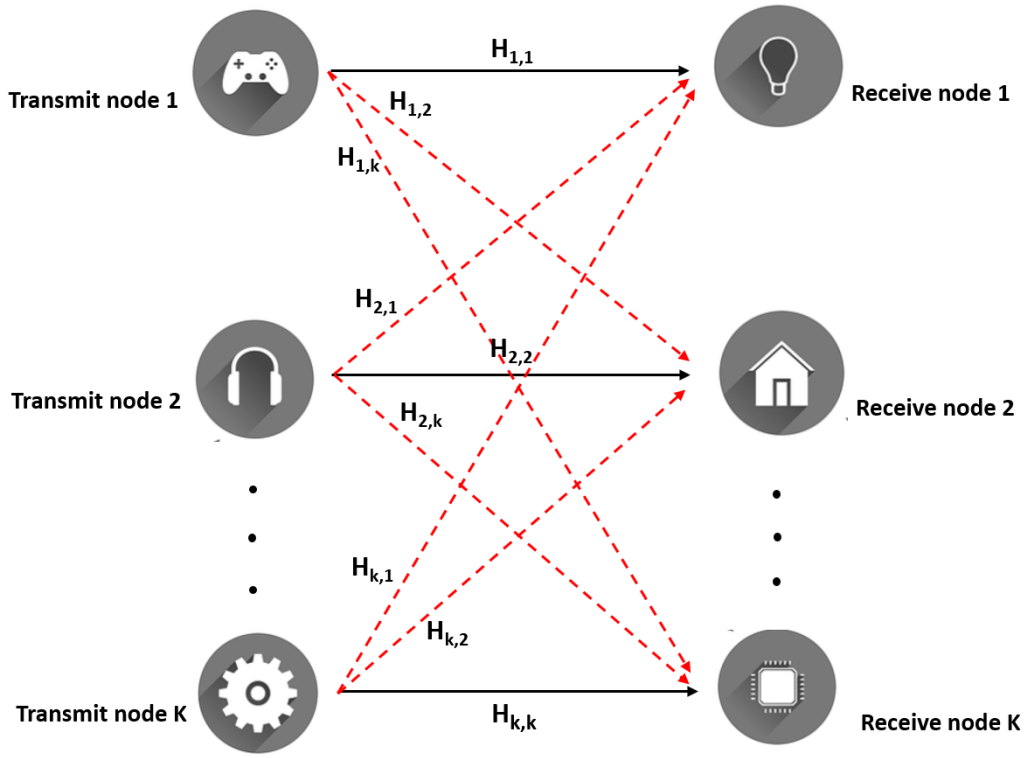


Figure 4.1: An illustration of a mmWave interference channel involving K pairs of IoT devices, where solid black lines represent direct links and the dashed red lines denote interfering links.

$$\mathbf{x}_j = \mathbf{V}_{j,\mathrm{RF}}\mathbf{V}_{j,\mathrm{BB}}\mathbf{s}_j, \tag{4.4}$$

where $\mathbf{s}_j \in \mathbb{C}^{N_\mathrm{s}\times 1}$ denotes the transmit data vector, which is assumed to be complex with zero mean and unit variance, i.e., $\mathbf{s}_j \sim \mathcal{CN}(0, \mathbf{I}_{N_\mathrm{s}})$, and $\|\mathbf{V}_{j,\mathrm{RF}}\mathbf{V}_{j,\mathrm{BB}}\|_F^2 \leq P_j$. Here, P_j represents the transmit power of the jth device.

The channel output at the receiving device j is given by

$$\mathbf{y}_j = \mathbf{H}_{j,j}\left(\mathbf{x}_j + \boldsymbol{\eta}_j\right) + \sum_{k\neq j}^{K} \mathbf{H}_{k,j}\left(\mathbf{x}_k + \boldsymbol{\eta}_k\right) + \boldsymbol{\xi}_j + \mathbf{z}_j, \tag{4.5}$$

where $\mathbf{y}_j \in \mathbb{C}^{N_\mathrm{R}\times 1}$ is the received signal, $\mathbf{x}_j \in \mathbb{C}^{N_\mathrm{T}\times 1}$ is the transmitted signal from device j and $\mathbf{x}_k \in \mathbb{C}^{N_\mathrm{T}\times 1}$ is the interference received from device k. Furthermore, $\mathbf{H}_{k,j} \in \mathbb{C}^{N_\mathrm{R}\times N_\mathrm{T}}$ represents the mmWave channel matrix from the transmitting device j to receiver k, and $\mathbf{z}_j \in \mathbb{C}^{N_\mathrm{R}\times 1}$ denotes the circularly symmetric AWGN with zero mean and variance σ_z^2. In addition, we have considered the hardware impairments which are unavoidable in practical mmWave transceivers in equation

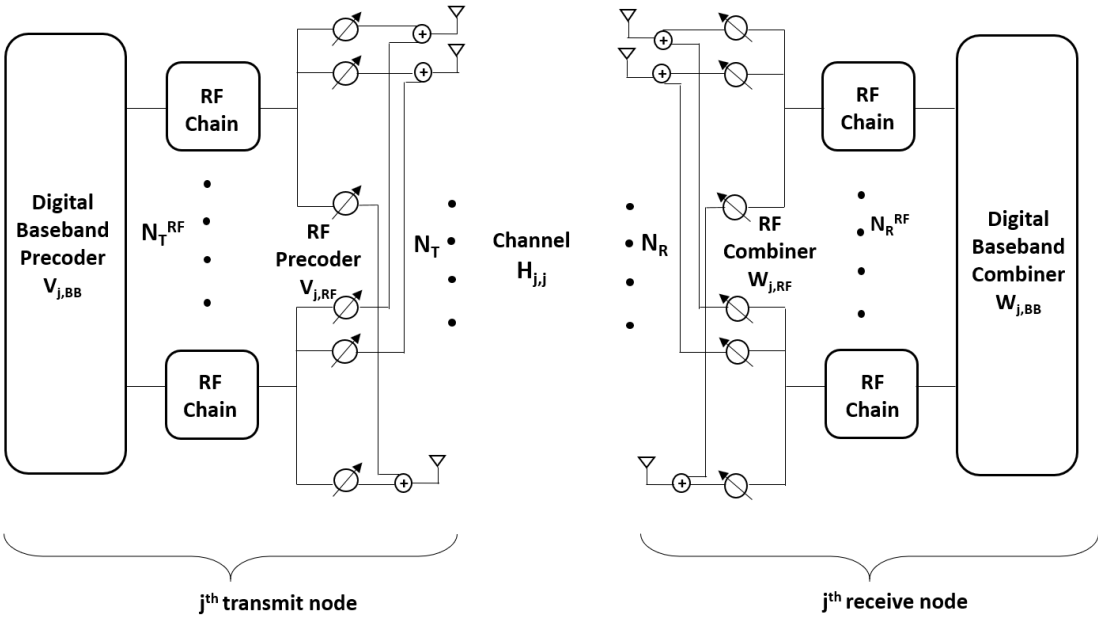


Figure 4.2: An illustration of hybrid processing between the jth transmit–receive pair.

(4.5). Accordingly, at the transmitter, $\boldsymbol{\eta}_j \in \mathbb{C}^{N_{\mathrm{T}} \times 1}$ denotes the transmit distortion, which encompasses the effects of non-linearities in the DAC, phase noise and additive power-amplifier noise. The covariance matrix of $\boldsymbol{\eta}_j$ is proportional to the energy of the desired signal $\mathbf{x}_j$. In particular, $\boldsymbol{\eta}_j$ is modelled as [103, 104]

$$\boldsymbol{\eta}_j \sim \mathcal{CN}\left(\mathbf{0}, \kappa \,\mathrm{diag}\left(\mathbf{V}_{j,\mathrm{RF}}\mathbf{V}_{j,\mathrm{BB}}\mathbf{V}_{j,\mathrm{BB}}^H\mathbf{V}_{j,\mathrm{RF}}^H\right)\right), \;\; \boldsymbol{\eta}_j \perp \mathbf{x}_j, \tag{4.6}$$

where $\perp$ denotes statistical independence, and $\kappa \ll 1$ is the level of impairment at the transmitter [105] .

Similarly, at the receiver, $\boldsymbol{\xi}_j$ represents receive distortions which approximate the combined effects of the non-linearities in the ADC, phase noise and additive gain control noise. The covariance matrix of $\boldsymbol{\xi}_j$ is given by β times the energy of the undistorted received signal at device j, with $\beta \ll 1$. Hence, $\boldsymbol{\xi}_j$ is modelled as

$$\boldsymbol{\xi}_j \sim \mathcal{CN}\left(\mathbf{0}, \beta \;\mathrm{diag}\left(\boldsymbol{\Upsilon}_j\right)\right), \;\; \boldsymbol{\xi}_j \perp \mathbf{u}_j, \tag{4.7}$$

where $\mathbf{u}_j$ is the undistorted received signal at device j, given as $\mathbf{u}_j = \mathbf{y}_j - \boldsymbol{\xi}_j$, and $\boldsymbol{\Upsilon}_j = \mathrm{cov}\{\mathbf{u}_j\}$ and β represents the level of impairment at the receiver.

We also assume that the receive devices are equipped with hybrid processors.

Hence, the received signal at device j is processed through an $N_{\mathrm{R}} \times N_{\mathrm{R}}^{\mathrm{RF}}$ RF combiner $\mathbf{W}_{j,\mathrm{RF}}$ followed by an $N_{\mathrm{R}}^{\mathrm{RF}} \times N_{\mathrm{s}}$ baseband combiner $\mathbf{W}_{j,\mathrm{BB}}$. Thus, the output signal at the receiving device j is given by

$$\begin{aligned}
\hat{\mathbf{s}}_j &= \mathbf{W}_{j,\mathrm{BB}}^H \mathbf{W}_{j,\mathrm{RF}}^H \mathbf{y}_j \\
&= \mathbf{W}_{j,\mathrm{BB}}^H \mathbf{W}_{j,\mathrm{RF}}^H \mathbf{H}_{j,j} \left(\mathbf{V}_{j,\mathrm{RF}} \mathbf{V}_{j,\mathrm{BB}} \mathbf{s}_j + \boldsymbol{\eta}_j \right) \qquad (4.8) \\
&\quad + \mathbf{W}_{j,\mathrm{BB}}^H \mathbf{W}_{j,\mathrm{RF}}^H \sum_{k \neq j}^{K} \mathbf{H}_{k,j} \left(\mathbf{V}_{k,\mathrm{RF}} \mathbf{V}_{k,\mathrm{BB}} \mathbf{s}_k + \boldsymbol{\eta}_k \right) + \mathbf{W}_{j,\mathrm{BB}}^H \mathbf{W}_{j,\mathrm{RF}}^H \left(\boldsymbol{\xi}_j + \mathbf{z}_j \right) \\
&= \mathbf{W}_{j,\mathrm{BB}}^H \mathbf{W}_{j,\mathrm{BB}}^H \mathbf{H}_{j,j} \mathbf{V}_{j,\mathrm{RF}} \mathbf{V}_{j,\mathrm{BB}} \mathbf{s}_j + \mathbf{m}_j, \qquad (4.9)
\end{aligned}$$

where $\mathbf{m}_j$ is the interference plus noise term, given as

$$\begin{aligned}
\mathbf{m}_j &= \mathbf{W}_{j,\mathrm{BB}}^H \mathbf{W}_{j,\mathrm{RF}}^H \mathbf{H}_{j,j} \mathbf{V}_{j,\mathrm{RF}} \mathbf{V}_{j,\mathrm{BB}} \boldsymbol{\eta}_j + \mathbf{W}_{j,\mathrm{BB}}^H \mathbf{W}_{j,\mathrm{RF}}^H \left(\boldsymbol{\xi}_j + \mathbf{z}_j \right) \\
&+ \mathbf{W}_{j,\mathrm{BB}}^H \mathbf{W}_{j,\mathrm{RF}}^H \sum_{k \neq j}^{K} \mathbf{H}_{k,j} \left(\mathbf{V}_{k,\mathrm{RF}} \mathbf{V}_{k,\mathrm{BB}} \, \mathbf{s}_k + \boldsymbol{\eta}_k \right). \qquad (4.10)
\end{aligned}$$

Considering the perfect CSI case, where perfect channel knowledge is available at the transmitting nodes, the achievable spectral efficiency of receive device j is given as

$$S_j = \log|\mathbf{I}_{N_{\mathrm{s}}} + \mathbf{W}_{j,\mathrm{BB}}^H \mathbf{W}_{j,\mathrm{RF}}^H \mathbf{H}_{j,j} \mathbf{V}_{j,\mathrm{RF}} \mathbf{V}_{j,\mathrm{BB}} \mathbf{V}_{j,\mathrm{BB}}^H \mathbf{V}_{j,\mathrm{RF}}^H \mathbf{H}_{j,j}^H \mathbf{W}_{j,\mathrm{RF}} \mathbf{W}_{j,\mathrm{BB}} \mathbf{R}_j^{-1}|, \qquad (4.11)$$

where $\mathbf{R}_j^{-1}$ is the covariance matrix of $\mathbf{m}_j$, given as[2]

$$\begin{aligned}
\mathbf{R}_j &= \kappa \mathbf{W}_{j,\mathrm{BB}}^H \mathbf{W}_{j,\mathrm{RF}}^H \mathbf{H}_{j,j} \mathrm{diag}\left(\mathbf{V}_{j,\mathrm{RF}} \mathbf{V}_{j,\mathrm{BB}} \mathbf{V}_{j,\mathrm{BB}}^H \mathbf{V}_{j,\mathrm{RF}}^H \right) \mathbf{H}_{j,j}^H \mathbf{W}_{j,\mathrm{RF}} \mathbf{W}_{j,\mathrm{BB}} \\
&+ \mathbf{W}_{j,\mathrm{BB}}^H \mathbf{W}_{j,\mathrm{RF}}^H \sum_{k \neq j}^{K} \left[\mathbf{H}_{k,j} \mathbf{V}_{k,\mathrm{RF}} \mathbf{V}_{k,\mathrm{BB}} \mathbf{V}_{k,\mathrm{BB}}^H \mathbf{V}_{k,\mathrm{RF}}^H \mathbf{H}_{k,j}^H \right] \mathbf{W}_{j,\mathrm{RF}} \mathbf{W}_{j,\mathrm{BB}} \\
&+ \kappa \mathbf{W}_{j,\mathrm{BB}}^H \mathbf{W}_{j,\mathrm{RF}}^H \sum_{k \neq j}^{K} \left[\mathbf{H}_{k,j} \mathrm{diag}\left(\mathbf{V}_{k,\mathrm{RF}} \mathbf{V}_{k,\mathrm{BB}} \mathbf{V}_{k,\mathrm{BB}}^H \mathbf{V}_{k,\mathrm{RF}}^H \right) \mathbf{H}_{k,j}^H \right] \mathbf{W}_{j,\mathrm{RF}} \mathbf{W}_{j,\mathrm{BB}} \\
&+ \beta \mathbf{W}_{j,\mathrm{BB}}^H \mathbf{W}_{j,\mathrm{RF}}^H \mathrm{diag}\left(\mathbf{H}_{j,j} \mathbf{V}_{j,\mathrm{RF}} \mathbf{V}_{j,\mathrm{BB}} \mathbf{V}_{j,\mathrm{BB}}^H \mathbf{V}_{j,\mathrm{RF}}^H \mathbf{H}_{j,j}^H \right) \mathbf{W}_{j,\mathrm{RF}} \mathbf{W}_{j,\mathrm{BB}} \\
&+ \beta \mathbf{W}_{j,\mathrm{BB}}^H \mathbf{W}_{j,\mathrm{RF}}^H \sum_{k \neq j}^{K} \left[\mathrm{diag}\left(\mathbf{H}_{k,j} \mathbf{V}_{k,\mathrm{RF}} \mathbf{V}_{k,\mathrm{BB}} \mathbf{V}_{k,\mathrm{BB}}^H \mathbf{V}_{k,\mathrm{RF}}^H \mathbf{H}_{k,j}^H \right) \right] \mathbf{W}_{j,\mathrm{RF}} \mathbf{W}_{j,\mathrm{BB}} \\
&+ \sigma_z^2 \mathbf{W}_{j,\mathrm{BB}}^H \mathbf{W}_{j,\mathrm{RF}}^H \mathbf{W}_{j,\mathrm{RF}} \mathbf{W}_{j,\mathrm{BB}}, \qquad (4.12)
\end{aligned}$$

[2]The terms including the multiplication of κ and β have been neglected in equation (4.12) since $\kappa \ll 1$ and $\beta \ll 1$.

and the corresponding sum spectral efficiency is given as

$$R = \sum_{j=1}^{K} S_j. \tag{4.13}$$

The imperfect CSI case will be considered in section 4.5.

4.2.1.1 Channel model

The mmWave propagation environment is characterised by limited scattering due to high free-space path-loss. Hence, a narrowband clustered model, which is based on the Saleh–Valenzuela model, where each scatterer contributes to a signal propagation path, can be accurately used to describe the mmWave channel [30]. Thus, the mmWave channel matrix $\mathbf{H}_{k,j}$ is expressed as

$$\mathbf{H}_{k,j} = \sqrt{\frac{N_{\mathrm{T}} N_{\mathrm{R}}}{N_{\mathrm{cl}} N_{\mathrm{p}}}} \sum_{l=1}^{N_{\mathrm{cl}}} \sum_{k=1}^{N_{\mathrm{p}}} \alpha_{l,k}^{(k,j)} \mathbf{a}_{\mathrm{R}}(\phi_{k,\mathrm{R}}^{l,k}, \theta_{k,\mathrm{R}}^{l,k}) \mathbf{a}_{\mathrm{T}}^{H}(\phi_{j,\mathrm{T}}^{l,k}, \theta_{j,\mathrm{T}}^{l,k}), \tag{4.14}$$

where N_{cl} and N_{p} denote the number of scattering clusters and number of paths per cluster, respectively, and $\alpha_{l,k}^{(k,j)}$ represents the complex gain of the kth path in the lth cluster. We assume $\alpha_{l,k}^{(k,j)}$ to be Rayleigh-distributed, i.e., $\alpha_{l,k}^{(k,j)} \sim \mathcal{CN}(0, \sigma_{\alpha,l}^2)$, where $\sigma_{\alpha,l}^2$ is the average power in the lth cluster. $\phi_{k,\mathrm{R}}^{l,k}$ and $\theta_{k,\mathrm{R}}^{l,k}$ denote the azimuth and elevation AoAs at receive device k, while $\phi_{j,\mathrm{T}}^{l,k}$, and $\theta_{j,\mathrm{T}}^{l,k}$ are the corresponding AoDs at transmit device j. Finally, $\mathbf{a}_{\mathrm{T}}$ and $\mathbf{a}_{\mathrm{R}}$ represent the transmit and receive antenna array response vectors, respectively.

We consider UPAs to model the array response vector which takes into account the azimuth and elevation angles and enables three-dimensional (3D) beamforming. In particular, at transmit device j, the array response vector is given by

$$\begin{aligned}\mathbf{a}_{\mathrm{T}}(\phi_{j,\mathrm{T}}, \theta_{j,\mathrm{T}}) = &\frac{1}{\sqrt{N_{\mathrm{T}}}} \Big[1, e^{i2\frac{\pi}{\lambda} d\left(a \sin(\phi_{j,\mathrm{T}}) \sin(\theta_{j,\mathrm{T}}) + b \cos(\theta_{j,\mathrm{T}})\right)}, \\ &\ldots e^{i2\frac{\pi}{\lambda} d\left((A-1) \sin(\phi_{j,\mathrm{T}}) \sin(\theta_{j,\mathrm{T}}) + (B-1) \cos(\theta_{j,\mathrm{T}})\right)}\Big],\end{aligned} \tag{4.15}$$

where λ denotes the wavelength, d represents the distance between antenna elements, and $0 \leq a \leq A$ and $0 \leq b \leq B$ are the indices of antenna elements in the 2D plane with $N_{\mathrm{T}} = AB$. The array response vector of the receive device k

follows a similar representation and is expressed as

$$\mathbf{a}_{\mathrm{R}}(\phi_{k,T},\theta_{k,T}) = \frac{1}{\sqrt{N_{\mathrm{R}}}}\Big[1, e^{i2\frac{\pi}{\lambda}d\left(a\sin(\phi_{k,\mathrm{R}})\sin(\theta_{k,\mathrm{R}})+b\cos(\theta_{k,\mathrm{R}})\right)}, \\ \ldots e^{i2\frac{\pi}{\lambda}d\left((A-1)\sin(\phi_{k,\mathrm{R}})\sin(\theta_{k,\mathrm{R}})+(B-1)\cos(\theta_{k,\mathrm{R}})\right)}\Big]. \tag{4.16}$$

4.2.1.2 Power consumption model

To calculate the EE effectively, we consider power consumption at the transmitter using the linear power model in [106]. The power consumption at the receiver side is neglected since in downlink transmissions, the power consumed by transmitter takes main possession [107, 108]. Thus, the power model at the transmitter side of a mmWave system can be given by

$$P_{\mathrm{total}} = \vartheta P_T + P_C, \tag{4.17}$$

where $\vartheta \geq 1$ denotes the inefficiency of the power amplifier and is dependent on the design and implementation of the power amplifier, and P_T is the transmitted power, which is expressed as

$$P_T = \sum_{j=1}^{K} \|\mathbf{V}_{j,\mathrm{RF}}\mathbf{V}_{j,\mathrm{BB}}\|_F^2. \tag{4.18}$$

The circuit power consumption P_C is independent of P_T and composed of two parts: the dynamic power P_{dyn} and the static power P_{sta}. P_{dyn} in a hybrid architecture describes both the power consumed by the phase shifters and the power radiation to the components of all RF chains in the circuit, e.g., mixers, local oscillators, DACs and frequency synthesizers. Accordingly,

$$P_{\mathrm{dyn}} = \sum_{j=1}^{K} N_{\mathrm{T}}^{j,\mathrm{RF}} P_{\mathrm{RF}} + N_{j,\mathrm{PS}} P_{\mathrm{PS}}, \tag{4.19}$$

where P_{RF} is the power consumed by components of the RF chain, which depends on the number of RF chains, and P_{PS} is the power consumed by the phase shifters i.e., the power required for excitation of the phase shifters and also to compensate for insertion losses [109]. The number of phase shifters for a fully connected hybrid architecture is given as $N_{\mathrm{PS}} = N_{\mathrm{T}} N_{\mathrm{T}}^{\mathrm{RF}}$. Furthermore, P_{sta} is the static circuit power independent of the number of transmit antennas and describes the

power consumed by the cooling system, power supply and baseband processors.

4.2.1.3 Problem formulation

The EE metric is defined as the ratio of the WSR and the total power consumption, and is expressed as

$$\mathrm{EE}(\mathbf{V}_{\mathrm{RF}}, \mathbf{V}_{\mathrm{BB}}, \mathbf{W}_{\mathrm{RF}}, \mathbf{W}_{\mathrm{BB}}) = \frac{\sum_{j=1}^{K} \varphi_j S_j(\mathbf{V}_{\mathrm{RF}}, \mathbf{V}_{\mathrm{BB}}, \mathbf{W}_{\mathrm{RF}}, \mathbf{W}_{\mathrm{BB}})}{\sum_{j=1}^{K} \left[\vartheta \|\mathbf{V}_{j,\mathrm{RF}} \mathbf{V}_{j,\mathrm{BB}}\|_F^2 + N_{\mathrm{T}}^{j,\mathrm{RF}} P_{\mathrm{RF}} + N_{j,\mathrm{PS}} P_{\mathrm{PS}} + P_{\mathrm{sta}} \right]}, \tag{4.20}$$

where φ_j is the weight used to denote the priority of device j in the system. Throughout this work, the EE is measured in bits/J/Hz.

Our analysis is concerned with the EE maximisation as a design criteria for obtaining the near-optimal hybrid precoders and combiners. Hence, the EE maximisation problem is formulated as

$$\max_{\mathbf{V}_{\mathrm{RF}}, \mathbf{V}_{\mathrm{BB}}, \mathbf{W}_{\mathrm{RF}}, \mathbf{W}_{\mathrm{BB}}} \quad \mathrm{EE}(\mathbf{V}_{\mathrm{RF}}, \mathbf{V}_{\mathrm{BB}}, \mathbf{W}_{\mathrm{RF}}, \mathbf{W}_{\mathrm{BB}}) \tag{4.21a}$$

$$\text{s.t.} \quad \mathbf{W}_{\mathrm{RF}} \in \mathcal{W}_{\mathrm{RF}} \tag{4.21b}$$

$$\mathbf{V}_{\mathrm{RF}} \in \mathcal{V}_{\mathrm{RF}} \tag{4.21c}$$

$$\|\mathbf{V}_{j,\mathrm{RF}} \mathbf{V}_{j,\mathrm{BB}}\|_F^2 \leq P_j, \ \forall\, j, \tag{4.21d}$$

where P_j is the transmit power constraint at device j, $\mathcal{W}_{\mathrm{RF}}$ and $\mathcal{V}_{\mathrm{RF}}$ represent the feasible RF combiner and precoder sets with constant-magnitude entries respectively.

4.3 EE Maximisation

In this section, we focus on designing the mmWave hybrid precoders $(\mathbf{V}_{\mathrm{RF}}, \mathbf{V}_{\mathrm{BB}})$ and combiners $(\mathbf{W}_{\mathrm{RF}}, \mathbf{W}_{\mathrm{BB}})$ in order to improve the network's EE by solving the problem described in equations (4.21a)–(4.21d). Note that the optimisation problem (4.21a)–(4.21d) is non-convex due to the coupling of optimisation variables and the non-convex constraints in equations (4.21c) and (4.21d). Hence, it is difficult to obtain a tractable optimal solution of the joint problem in its

current state. To this end, we will first transform the problem into an equivalent univariate joint optimisation problem and then, employ an alternating maximisation algorithm. Afterwards, the near-optimal mmWave hybrid precoders will be achieved by minimising the Euclidean distance between the optimal univariate precoder and the multivariate precoders. Finally, the hybrid combiners will be obtained by the minimisation of the minimum mean squared error (MMSE) between the transmitted and received signals.

4.3.1 Univariate EE maximisation

The equivalent univariate EE maximisation problem is computed by letting the product of the hybrid precoders $\mathbf{V}_{\mathrm{RF}}\mathbf{V}_{\mathrm{BB}} = \mathbf{V}$, where $\mathbf{V} \in \mathbb{C}^{N_{\mathrm{T}} \times N_{\mathrm{s}}}$, and the multiplication of the hybrid combiners $\mathbf{W}_{\mathrm{RF}}\mathbf{W}_{\mathrm{BB}} = \mathbf{W}$, where $\mathbf{W} \in \mathbb{C}^{N_{\mathrm{R}} \times N_{\mathrm{s}}}$. Accordingly, the problem (4.21a)–(4.21d) can be reformulated by assuming fixed analog precoders and combiners [82] and expressed as

$$\max_{\mathbf{V},\mathbf{W}} \ \mathrm{EE}(\mathbf{V}, \mathbf{W}) \tag{4.22a}$$

$$\text{s.t. } \mathrm{tr}\left\{\mathbf{V}_j \mathbf{V}_j^H\right\} \leq P_j, \ \forall\, j, \tag{4.22b}$$

where $\mathrm{EE}(\mathbf{V}, \mathbf{W})$ is expressed as

$$\mathrm{EE}(\mathbf{V}, \mathbf{W}) = \frac{\sum\limits_{j=1}^{K} \varphi_j \log\left|\mathbf{I}_{N_{\mathrm{s}}} + \mathbf{W}_j^H \mathbf{H}_{j,j} \mathbf{V}_j \mathbf{V}_j^H \mathbf{H}_{j,j}^H \mathbf{W}_j \mathbf{R}_{n,j}^{-1}\right|}{\sum\limits_{j=1}^{K} \left[\vartheta \mathrm{tr}\left\{\mathbf{V}_j \mathbf{V}_j^H\right\} + N_{\mathrm{T}}^{j,\mathrm{RF}} P_{\mathrm{RF}} + N_{j,\mathrm{PS}} P_{\mathrm{PS}} + P_{\mathrm{sta}}\right]}, \tag{4.23}$$

where $\mathbf{R}_{n,j}$ is given as

$$\begin{aligned}
\mathbf{R}_{n,j} \ \approx \ & \kappa \mathbf{W}_j^H \mathbf{H}_{j,j} \mathrm{diag}\left(\mathbf{V}_j \mathbf{V}_j^H\right) \mathbf{H}_{j,j}^H \mathbf{W}_j + \mathbf{W}_j^H \sum_{k \neq j}^{K} \left[\mathbf{H}_{k,j} \mathbf{V}_k \mathbf{V}_k^H \mathbf{H}_{k,j}^H\right] \mathbf{W}_j \\
+ \ & \kappa \mathbf{W}_j^H \sum_{k \neq j}^{K} \left[\mathbf{H}_{k,j} \mathrm{diag}\left(\mathbf{V}_k \mathbf{V}_k^H\right) \mathbf{H}_{k,j}^H\right] \mathbf{W}_j + \beta \mathbf{W}_j^H \mathrm{diag}\left(\mathbf{H}_{j,j} \mathbf{V}_j \mathbf{V}_j^H \mathbf{H}_{j,j}^H\right) \mathbf{W}_j \\
+ \ & \beta \mathbf{W}_j^H \sum_{k \neq j}^{K} \left[\mathrm{diag}\left(\mathbf{H}_{k,j} \mathbf{V}_k \mathbf{V}_k^H \mathbf{H}_{k,j}^H\right)\right] \mathbf{W}_j + \sigma_z^2 \mathbf{W}_j^H \mathbf{W}_j.
\end{aligned} \tag{4.24}$$

We note that the transformed EE maximisation problem (4.22a)–(4.22b) is not jointly convex over the precoding matrix $\mathbf{V}$ and combining matrix $\mathbf{W}$ but it is

component-wise convex over $\mathbf{V}$ and $\mathbf{W}$. Consequently, we will employ an iterative algorithm method to obtain the efficient solutions of $\mathbf{V}$ and $\mathbf{W}$. In particular, we update the precoding matrix $\mathbf{V}$ when the combining matrix $\mathbf{W}$ is fixed, and then, with $\mathbf{V}$ obtained from the previous step, we update the combining matrix $\mathbf{W}$. This alternating maximisation process continues until convergence or a predefined number of iterations is reached. Under a fixed receive combining matrix $\mathbf{W}$, the EE is given by

$$\begin{aligned} \mathrm{EE}(\mathbf{V}) &= \frac{f_1(\mathbf{V})}{f_2(\mathbf{V})}, \\ &= \frac{\sum_{j=1}^{K} \varphi_j \mathcal{I}_j(\mathbf{V})}{\sum_{j=1}^{K} \left[\vartheta \mathrm{tr}\left\{ \mathbf{V}_j \mathbf{V}_j^H \right\} + N_{\mathrm{T}}^{j,\mathrm{RF}} P_{\mathrm{RF}} + N_{j,\mathrm{PS}} P_{\mathrm{PS}} + P_{\mathrm{sta}} \right]}, \end{aligned} \tag{4.25}$$

where $\mathcal{I}_j = \log \left| \mathbf{I}_{N_{\mathrm{R}}} + \mathbf{\Sigma}_j^{-1} \mathbf{H}_{j,j} \mathbf{V}_j \mathbf{V}_j^H \mathbf{H}_{j,j}^H \right|$ is the rate, and $\mathbf{\Sigma}_j$ is the covariance matrix of the interference-plus-noise terms, given as

$$\begin{aligned} \mathbf{\Sigma}_j &\approx \kappa \mathbf{H}_{j,j} \mathrm{diag}\left(\mathbf{V}_j \mathbf{V}_j^H \right) \mathbf{H}_{j,j}^H + \sum_{k \neq j}^{K} \mathbf{H}_{k,j} \left(\mathbf{V}_k \mathbf{V}_k^H + \kappa \, \mathrm{diag}\left(\mathbf{V}_k \mathbf{V}_k^H \right) \right) \mathbf{H}_{k,j}^H \\ &+ \beta \, \mathrm{diag}\left(\mathbf{H}_{j,j} \mathbf{V}_j \mathbf{V}_j^H \mathbf{H}_{j,j}^H \right) + \sum_{k \neq j}^{K} \beta \, \left(\mathrm{diag}\left(\mathbf{H}_{k,j} \mathbf{V}_k \mathbf{V}_k^H \mathbf{H}_{k,j}^H \right) \right) + \sigma_z^2 \mathbf{I}_{N_{\mathrm{R}}}. \end{aligned} \tag{4.26}$$

Accordingly, the EE maximisation problem to compute the optimum precoding matrix can be formulated as

$$\max_{\mathbf{V}} \ \mathrm{EE}(\mathbf{V}) \tag{4.27a}$$

$$\text{s.t. } \mathrm{tr}\left\{ \mathbf{V}_j \mathbf{V}_j^H \right\} \leq P_j, \ \forall \, j, \tag{4.27b}$$

Nonetheless, the optimisation problem described by equations (4.27a)–(4.27b) is still non-convex due to the fractional form of the objective function in equation (4.25). Thus, we will exploit the relationship between the fractional and parametric programming problems to transform the fractional objective function into its non-fractional equivalent [110].

Lemma 4.1. *Let the maximum EE be denoted as g^*. Then, the optimal transmit*

precoding matrix $\mathbf{V}^{opt}$ *can achieve* $g^* = \frac{f_1(\mathbf{V}^{opt})}{f_2(\mathbf{V}^{opt})}$ *if and only if*

$$\begin{aligned} \max_{\mathbf{V}} \quad & f_1(\mathbf{V}) - g^* f_2(\mathbf{V}) \\ & = f_1(\mathbf{V}^{opt}) - g^* f_2(\mathbf{V}^{opt}) = 0, \end{aligned} \tag{4.28}$$

provided that $f_1(\mathbf{V}) \geq 0$ *and* $f_2(\mathbf{V}) > 0$.

Proof. The maximum EE can be expressed as

$$g^* = \frac{f_1(\mathbf{V}^{\text{opt}})}{f_2(\mathbf{V}^{\text{opt}})} \rightarrow f_1(\mathbf{V}^{\text{opt}}) - g^* f_2(\mathbf{V}^{\text{opt}}) = 0 \text{ and,} \tag{4.29}$$

$$g^* \geq \frac{f_1(\mathbf{V})}{f_2(\mathbf{V})} \rightarrow f_1(\mathbf{V}) - g^* f_2(\mathbf{V}) \leq 0. \tag{4.30}$$

From equation (4.29) and (4.30), we can conclude that $\max_{\mathbf{V}} f_1(\mathbf{V}) - g^* f_2(\mathbf{V}) = 0$ when $\mathbf{V} = \mathbf{V}^{\text{opt}}$.

Hence, it can be deduced from Lemma 4.1 that any fractional objective function can be transformed into an equivalent function in subtractive form. Within this context, for a given g, we rewrite the problem (4.27a)–(4.27b) as

$$\max_{\mathbf{V}} \ \sum_{j=1}^{K} \varphi_j \mathcal{I}_j(\mathbf{V}) - g f_2(\mathbf{V}) \tag{4.31a}$$

$$\text{s.t. } \text{tr}\left\{\mathbf{V}_j \mathbf{V}_j^H\right\} \leq P_j, \ \forall\, j. \tag{4.31b}$$

Here, g can be treated as the EE parameter depicting the severity of high power consumption, and the optimal solution of the problem (4.27a)–(4.27b) is achieved when we determine the parameter g such that non-fractional problem (4.31a)–(4.31b) is zero. Interestingly, if the value of g is zero then the transformed optimisation problem reduces to the WSR maximisation problem since $\sum_{j=1}^{K} \varphi_j \mathcal{I}_j(\mathbf{V})$ represents the WSR.

Despite the non-fractional transformation of (4.27a)–(4.27b), the resulting optimisation problem (4.31a)–(4.31b) is still non-convex due to the coupling of the variable $\mathbf{V}$ in $\mathcal{I}_j(\mathbf{V})$ and $f_2(\mathbf{V})$. To tackle this, we exploit the relationship between the WSR and WMMSE [111, 112]. We first need to define the MSE which will be employed in determining the relationship between WSR and WMMSE.

Thus, the MSE matrix of the jth transmitter–receiver pair is given by

$$\mathbf{M}_j = \left(\mathbf{W}_j^H \mathbf{y}_j - \mathbf{I}_{N_\mathrm{s}}\right)\left(\mathbf{W}_j^H \mathbf{y}_j - \mathbf{I}_{N_\mathrm{s}}\right)^H + \mathbf{W}_j \mathbf{\Sigma}_j \mathbf{W}_j^H, \tag{4.32}$$

where $\mathbf{y}_j$ and $\mathbf{\Sigma}_j$ have been previously defined in equations (4.5) and (4.26), respectively. The optimal receiver at device j is the MMSE receiver, which is expressed as

$$\mathbf{W}_j = \mathbf{V}_j^H \mathbf{H}_{j,j}^H \left(\mathbf{H}_{j,j}\mathbf{V}_j\mathbf{V}_j^H\mathbf{H}_{j,j}^H + \mathbf{\Sigma}_j\right)^{-1}. \tag{4.33}$$

Given the independence of the input signals and noises, and using an argument parallel to that in [111, 112], we can establish an equivalence between the WSR in equation (4.31a) and a corresponding MMSE one. Accordingly, the WMMSE optimisation problem is formulated as

$$\max_{\mathbf{V},\mathbf{W},\mathbf{Q}} \; f_1(\mathbf{V},\mathbf{W},\mathbf{Q}) - g f_2(\mathbf{V}) \tag{4.34a}$$

$$\text{s.t.} \quad \mathrm{tr}\left\{\mathbf{V}_j\mathbf{V}_j^H\right\} \leq P_j, \; \forall\, j, \tag{4.34b}$$

where

$$f_1(\mathbf{V},\mathbf{W},\mathbf{Q}) = \sum_{j=1}^{K}\left[\mathrm{tr}\left\{\mathbf{Q}_j\mathbf{M}_j\right\} - \varphi_j \log\left|\frac{\ln 2}{\varphi_j}\mathbf{Q}_j\right| - \frac{N_\mathrm{s}\varphi_j}{\ln 2}\right], \tag{4.35}$$

in which $\mathbf{Q}_j$ represents the MSE weight matrix associated with device j.

Theorem 4.1. *The WMMSE problem described in equations* (4.34a)–(4.34b) *is equivalent to the WSR problem in equations* (4.31a)–(4.31b)*, since the solution for the precoders obtained from solving both problems are identical.*

Proof. The MSE matrix, $\mathbf{M}_j$, can be obtained by substituting the receiver $\mathbf{W}_j$ from equation (4.33) in equation (4.32). Next, considering equations (4.34a)–(4.34b), it can be seen that for fixed $\mathbf{V}$ and $\mathbf{W}$ the weight matrix $\mathbf{Q}_j$ is obtained as

$$\mathbf{Q}_j = \frac{\varphi_j}{\ln 2}\Big(\mathbf{E}_j(\mathbf{V})\Big)^{-1}, \tag{4.36}$$

where $\mathbf{E}_j$ is the MSE matrix is given by

$$\begin{aligned}\mathbf{E}_j &= \mathbb{E}\left\{\left(\mathbf{W}_j^H\mathbf{y}_j - \mathbf{s}_j\right)\left(\mathbf{W}_j^H\mathbf{y}_j - \mathbf{s}_j\right)^H\right\}, \\ &= \left(\mathbf{I}_{N_\mathrm{s}} + \mathbf{V}_j^H\mathbf{H}_{j,j}^H\mathbf{\Sigma}_j^{-1}\mathbf{H}_{j,j}\mathbf{V}_j\right).\end{aligned} \tag{4.37}$$

Now, substituting for $\mathbf{M}_j$ and $\mathbf{Q}_j$ in equation (4.34a) and using the relationship $\mathcal{I}_j(\mathbf{V}) = \log\left|\left(\mathbf{E}_j(\mathbf{V})\right)^{-1}\right|$, we obtain the WSR objective function in equation (4.31a).

Note that the EE maximisation problem described in equations (4.34a)–(4.34b) is not jointly convex in $\mathbf{V}$, $\mathbf{W}$ or $\mathbf{Q}$, but is separately convex for each variable. Having obtained the optimal solutions for $\mathbf{W}$ and $\mathbf{Q}$ in equations (4.33) and (4.36), respectively, we focus on determining the optimal solution of the precoder $\mathbf{V}$.

For fixed $\mathbf{Q}$ and $\mathbf{W}$, the optimisation problem to obtain $\mathbf{V}$ is expressed as

$$\max_{\mathbf{V}} \sum_{j=1}^{K} \left[-\mathrm{tr}\left\{\mathbf{Q}_j\mathbf{M}_j\right\} - g\vartheta \mathrm{tr}\left\{\mathbf{V}_j\mathbf{V}_j^H\right\}\right] \tag{4.38a}$$

$$\text{s.t. } \mathrm{tr}\left\{\mathbf{V}_j\mathbf{V}_j^H\right\} \leq P_j, \ \forall\, j. \tag{4.38b}$$

The Lagrange objective function of the problem (4.38a)–(4.38b) is given as

$$\mathcal{L} = \sum_{j=1}^{K} \left[-\mathrm{tr}\left\{\mathbf{Q}_j\mathbf{M}_j\right\} - g\vartheta \mathrm{tr}\left\{\mathbf{V}_j\mathbf{V}_j^H\right\}\right] - \sum_{j=1}^{K} \lambda_j \left[\mathrm{tr}\left\{\mathbf{V}_j\mathbf{V}_j^H\right\} - P_j\right], \tag{4.39}$$

where $\lambda_j \geq 0$ is the Lagrange multiplier associated with the transmit power constraint. By setting $\frac{\partial \mathcal{L}}{\partial \mathbf{V}_j^H} = 0$, we obtain the closed-form solution of the optimal precoder as

$$\mathbf{V}_j = \left(\lambda_j \mathbf{I}_{N_\mathrm{T}} + g\vartheta \mathbf{I}_{N_\mathrm{T}} + \mathbf{\Phi}_j\right)^{-1} \mathbf{H}_{j,j}^H \mathbf{W}_j^H \mathbf{Q}_j, \tag{4.40}$$

where $\mathbf{\Phi}_j$ is defined as

$$\begin{aligned} \mathbf{\Phi}_j &= \mathbf{H}_{j,j}^H \mathbf{W}_j^H \mathbf{Q}_j \mathbf{W}_j \mathbf{H}_{j,j} + \kappa \mathrm{diag}\left(\mathbf{H}_{j,j}^H \mathbf{W}_j^H \mathbf{Q}_j \mathbf{W}_j \mathbf{H}_{j,j}\right) \\ &+ \beta\, \mathbf{H}_{j,j}^H \mathrm{diag}\left(\mathbf{W}_j^H \mathbf{Q}_j \mathbf{W}_j\right) \mathbf{H}_{j,j}. \end{aligned} \tag{4.41}$$

The Lagrange multiplier in equation (4.40) should either be positive or zero and must satisfy

$$\mathrm{tr}\left\{\mathbf{V}_j\mathbf{V}_j^H\right\} = \sum_{l=1}^{N_\mathrm{T}} \frac{[\mathbf{B}_j]_l}{([\mathbf{D}_j]_l + \lambda_j)} = P_j, \tag{4.42}$$

where $\mathbf{B}_j = \mathbf{U}_j^H \mathbf{H}_{j,j}^H \mathbf{W}_j^H \mathbf{Q}_j \mathbf{Q}_j^H \mathbf{W}_j \mathbf{H}_{j,j} \mathbf{U}_j$, and $\mathbf{D}_j$ is computed from the eigen-

value decomposition of

$$(g\vartheta\mathbf{I}_{N_{\mathrm{T}}} + \mathbf{\Phi}_j) = \mathbf{U}_j\mathbf{D}_j\mathbf{U}_j^H. \tag{4.43}$$

Further, $[\mathbf{B}_j]_l$ denotes the lth element along the diagonal of $\mathbf{B}_j$. Hence, we numerically solve for λ_j by using linear search techniques [111].

Algorithm 4.1: WMMSE-based EE maximisation algorithm.

1 Set the iteration number $n = 0$ and Initialise $\mathbf{V}_j$ such that $\mathrm{tr}\left\{\mathbf{V}_j\mathbf{V}_j^H\right\} \leq P_j \ \forall\, j$
2 Set the EE parameter $g = 0$.
3 $n \leftarrow n + 1$ Update $\mathbf{W}_j \ \forall\, j$ by solving (4.33).
4 Update $\mathbf{Q}_j \ \forall\, j$ using (4.36).
5 Compute $\mathbf{V}_j \ \forall\, j$ by solving (4.34a)–(4.34b) .
6 Repeat steps 3–5 until convergence or a pre-defined number of iterates is reached.
7 **if** $f_1(\mathbf{V}, \mathbf{W}, \mathbf{Q}) - gf_2(\mathbf{V}) \leq \delta$ **then**
8 Stop the iterations.
9 **else**
10 $g = \frac{f_1(\mathbf{V},\mathbf{W},\mathbf{Q})}{f_2(\mathbf{V})}$, and go to step 3.
11 **end**
12 **end**

4.3.2 Discussion

4.3.2.1 Convergence

The WMMSE-based alternating process used to solve the EE maximisation problem is presented in Algorithm 4.1 and should be run for every coherence time. The algorithm increases the objective function monotonically at each step of the iterations. In particular, the updates of $\mathbf{W}, \mathbf{Q}, \mathbf{V}$ in steps 3–5 maximise the objective function (4.34a). As (4.34a) is upper-bounded by the transmit power constraint, the algorithm is guaranteed to converge. In addition, given that (4.34a) is differentiable, the convergence of equations (4.34a)–(4.34b) to a stationary point can be proven using the dual-coordinate ascent method [113].

4.3.2.2 Complexity analysis

In this subsection, we evaluate the complexity of the WMMSE-based EE maximisation algorithm by counting the number of floating point operations (flops) in each iteration. A flop is counted when either a complex summation or a complex multiplication occurs. The complexity of standard matrix operations relevant to our analysis are given as follows. Let $\mathbf{R} \in \mathcal{C}^{M\times N}$ and $\mathbf{S} \in \mathcal{C}^{N\times N}$ be the matrices to be utilised, then the complexity involved in the following computations are given as [114]

(1) matrix–matrix multiplication: $\mathbf{R} \times \mathbf{S}$ requires $2MN^2 - MN$ flops,

(2) matrix inversion: $\mathbf{S}^{-1}$ requires $N^3 + N^2 + N$ flops,

(3) matrix Hermitian–matrix multiplication: $\mathbf{R}^H\mathbf{R}$ requires $M(N^2+N)-\frac{N^2+N}{2}$ flops,

(4) diagonal matrix multiplication: $\mathbf{R}\,\mathrm{diag}\,(\mathbf{S})$ requires MN flops,

(5) scalar–matrix multiplication: $c\,\mathbf{R}$ requires MN flops.

Accordingly, the computational complexity involved in the evaluation of Algorithm 4.1 is given as follows.

- The complexity involved in computing the receiver, i.e.,

$$\mathbf{W}_j = \mathbf{V}_j^H\mathbf{H}_{j,j}^H\left(\mathbf{H}_{j,j}\mathbf{V}_j\mathbf{V}_j^H\mathbf{H}_{j,j}^H + \mathbf{\Sigma}_j\right)^{-1}$$

  is given as follows. The inverse function requires $N_{\mathrm{R}}^3 + N_{\mathrm{R}}^2 + N_{\mathrm{R}}$ flops, multiplication of the terms outside the inverse function requires $2N_{\mathrm{s}}N_{\mathrm{R}}^2 + 2N_{\mathrm{s}}N_{\mathrm{T}}N_{\mathrm{R}} - 2N_{\mathrm{s}}N_{\mathrm{R}}$ flops and multiplication of terms inside the inverse function requires $2N_{\mathrm{T}}\left(N_{\mathrm{s}}N_{\mathrm{T}} + N_{\mathrm{s}} - \frac{N_{\mathrm{T}}+1}{2} + 2N_{\mathrm{T}}N_{\mathrm{R}} + 2N_{\mathrm{R}}^2 - 2N_{\mathrm{R}}\right) - 2N_{\mathrm{R}}^2$ flops. Therefore, $\mathcal{O}\left(K(N_{\mathrm{R}}^3 + 2N_{\mathrm{T}}^2(N_{\mathrm{R}} + N_{\mathrm{s}}) + 4N_{\mathrm{R}}^2N_{\mathrm{T}})\right)$ flops are required to compute the optimal receive matrix $\mathbf{W}$.

- For the weight matrix, $\mathbf{Q}_j = \frac{\varphi_j}{\ln 2}\Big(\mathbf{E}_j(\mathbf{V})\Big)^{-1}$, the interference-plus-noise-matrix, $\mathbf{\Sigma}_j$, is already available since it was used in the calculation of $\mathbf{W}_j$. Thus, we only require $N_{\mathrm{R}}^3+N_{\mathrm{R}}^2+N_{\mathrm{R}}$ flops for the calculation of the inverse of $\mathbf{E}_j(\mathbf{V})$, and $2N_{\mathrm{s}}N_{\mathrm{R}}\Big(N_{\mathrm{s}} + N_{\mathrm{T}} + N_{\mathrm{R}} - 1 - \frac{N_{\mathrm{s}}}{2N_{\mathrm{R}}}\Big)$ flops for the multiplication in $\mathbf{E}_j(\mathbf{V})$. Thus, $\mathcal{O}\left(K(N_{\mathrm{R}}^3 + 2N_{\mathrm{R}}^2N_{\mathrm{s}} + 2N_{\mathrm{s}}^2N_{\mathrm{R}})\right)$ flops are required to calculate the optimal weight matrix.

- For the precoder $\mathbf{V}_j = (\lambda_j \mathbf{I}_{N_\mathrm{T}} + g\vartheta \mathbf{I}_{N_\mathrm{T}} + \mathbf{\Phi}_j)^{-1} \mathbf{H}_{j,j}^H \mathbf{W}_j^H \mathbf{Q}_j$, the inverse function requires $N_\mathrm{T}^3 + N_\mathrm{T}^2 + N_\mathrm{T}$ flops, $2N_\mathrm{s}^2 N_\mathrm{R} + 2N_\mathrm{s} N_\mathrm{R}^2 - 2N_\mathrm{R} N_\mathrm{s} - 2N_\mathrm{R}^2 + 4N_\mathrm{T} N_\mathrm{R}^2 + 4N_\mathrm{T}^2 N_\mathrm{R} - 2N_\mathrm{R} N_\mathrm{T} - 2N_\mathrm{T}^2$ flops are required to compute $\mathbf{\Phi}_j$ and $2N_\mathrm{s} N_\mathrm{T} (N_\mathrm{T} + N_\mathrm{R} - 1)$ flops are required for multiplying the inverse function with the terms outside the inverse. Hence, the total number of flops required for deriving the optimal $\mathbf{V}$ is $\mathcal{O}\left(K(N_\mathrm{T}^3 + 4N_\mathrm{T}^2 N_\mathrm{R} + 4N_\mathrm{R}^2 N_\mathrm{T})\right)$.

Remark 4.1. *Note that although the complexity of the proposed algorithm to obtain the energy-efficient precoder* $\mathbf{V}$ *is cubic in the number of transmit and receive antennas, the application for communication between IoT devices puts a contraint on the number of transmit and receive antennas to be used in each device and as such the moderate complexity permits the use of the algorithm in this case.*

4.4 EE Hybrid Transceiver Design

In mmWave communications, due to the hybrid nature of the transceivers, the EE univariate precoders and decoders obtained from the analysis in section 4.3 cannot be directly applied to the system. Therefore, in this section, we present the derivation of the energy-efficient hybrid precoders $(\mathbf{V}_\mathrm{RF}, \mathbf{V}_\mathrm{BB})$ for the transmitting devices and the corresponding hybrid combiners $(\mathbf{W}_\mathrm{RF}, \mathbf{W}_\mathrm{BB})$ for the receiving devices.

4.4.1 Design of hybrid precoders

The proposed hybrid precoders for the jth transmit device are designed by the minimisation of the Euclidean distance between the optimal energy-efficient precoder $\mathbf{V}_j^\mathrm{opt}$ in equation (4.40) and the product of $\mathbf{V}_{j,\mathrm{RF}}$ and $\mathbf{V}_{j,\mathrm{BB}}$. Using an argument parallel to that of [28, 115], we can establish that the hybrid precoders can be obtained by finding the projection of $\mathbf{V}_j^\mathrm{opt}$ on to the hybrid precoders $\mathbf{V}_{j,\mathrm{RF}}$ and $\mathbf{V}_{j,\mathrm{BB}}$, with $\mathbf{V}_{j,\mathrm{RF}}$ belonging to a feasible set of constant magnitude entries. Therefore, we consider the following optimisation problem

$$\min_{\mathbf{V}_\mathrm{RF}, \mathbf{V}_\mathrm{BB}} \|\mathbf{V}_j^\mathrm{opt} - \mathbf{V}_{j,\mathrm{RF}} \mathbf{V}_{j,\mathrm{BB}}\|_F \tag{4.44a}$$

$$\text{s.t.} \quad \mathbf{V}_{j,\mathrm{RF}} \in \mathcal{V}_{j,\mathrm{RF}}, \tag{4.44b}$$

$$\|\mathbf{V}_{j,\mathrm{RF}} \mathbf{V}_{j,\mathrm{BB}}\|_F^2 \leq P_j. \tag{4.44c}$$

The problem described by equations (4.44a)–(4.44c) is non-convex due to the constraint (4.44b) and it cannot be solved by standard optimisation tools. Accordingly, to deal with the non-convexity of equation(4.44b), we will follow the method in [115] which chooses vectors from a set of candidate beamformers whose elements can form the bases of $\mathbf{V}^{\text{opt}}$. If we choose the array response beamformers which are implemented by phase shifters, the problem (4.44a)–(4.44c) can be reformulated as

$$\min_{\mathbf{V}_{\text{RF}},\mathbf{V}_{\text{BB}}} \|\mathbf{V}_j^{\text{opt}} - \mathbf{V}_{j,\text{RF}}\mathbf{V}_{j,\text{BB}}\|_F \tag{4.45a}$$

$$\text{s.t.} \quad [\mathbf{V}_{j,\text{RF}}]_l \in \mathbf{a}_{\text{T}}(\phi_{j,\text{T}}^{l,k}, \theta_{j,\text{T}}^{l,k}) \; \forall\, l,k, \tag{4.45b}$$

$$\|\mathbf{V}_{j,\text{RF}}\mathbf{V}_{j,\text{BB}}\|_F^2 \leq P_j. \tag{4.45c}$$

Next, we can insert the constraint (4.45b) directly into the objective function (4.45a) to obtain the following equivalent problem as

$$\min_{\mathbf{V}_{\text{BB}}} \|\mathbf{V}_j^{\text{opt}} - \mathbf{A}_{j,\text{T}}\mathbf{V}_{j,\text{BB}}\|_F \tag{4.46a}$$

$$\text{s.t.} \|\mathbf{A}_{j,\text{T}}\mathbf{V}_{j,\text{BB}}\|_F^2 \leq P_j. \tag{4.46b}$$

$$\|\text{diag}\left(\mathbf{V}_{j,\text{BB}}\mathbf{V}_{j,\text{BB}}^H\right)\|_0 \leq N_{\text{T}}^{\text{RF}}, \tag{4.46c}$$

where $\mathbf{A}_{j,\text{T}} = [\mathbf{a}_{\text{T}}(\phi_{j,\text{T}}^{1,1}, \theta_{j,\text{T}}^{1,1}), \ldots, \mathbf{a}_{\text{T}}(\phi_{j,\text{T}}^{N_{\text{cl}},N_{\text{P}}}, \theta_{j,\text{T}}^{N_{\text{cl}},N_{\text{P}}})]$ is the array response matrix with AoDs pointing to the desired user, and equation (4.46c) is the sparsity

Algorithm 4.2: OMP-based algorithm for energy-efficient hybrid precoders.

Initialise $\mathbf{V}_{j,\text{RF}}$ and $\mathbf{V}_{j,\text{BB}}$.
Set $\mathbf{V}_{j,\text{res}} = \mathbf{V}_j^{\text{opt}}$ using (4.40).
for $i \leq N_{\text{T}}^{\text{RF}}$ **do**
  $\mathbf{\Omega}_j = \mathbf{A}_{j,\text{T}}^H \mathbf{V}_{j,\text{res}}$
  $[\text{value}, \text{index}] = \max\left(\text{diag}\left(\mathbf{\Omega}_j\mathbf{\Omega}_j^H\right)\right)$
  $k = \text{index}$
  $\mathbf{V}_{j,\text{RF}} = \left[\mathbf{V}_{j,\text{RF}} | \mathbf{A}_{j,\text{T}}^{\text{k}}\right]$
  $\mathbf{V}_{j,\text{BB}} = \left(\mathbf{V}_{j,\text{RF}}^H\mathbf{V}_{j,\text{RF}}\right)^{-1}\mathbf{V}_{j,\text{RF}}^H\mathbf{V}_{j,\text{res}}$
  Update $\mathbf{V}_{j,\text{res}} = \frac{\mathbf{V}_j^{\text{opt}} - \mathbf{V}_{j,\text{RF}}\mathbf{V}_{j,\text{BB}}}{\|\mathbf{V}_j^{\text{opt}} - \mathbf{V}_{j,\text{RF}}\mathbf{V}_{j,\text{BB}}\|_F}$
end
$\mathbf{V}_{j,\text{BB}} = \frac{\mathbf{V}_{j,\text{BB}}}{\mathbf{V}_{j,\text{RF}}\mathbf{V}_{j,\text{BB}}}\|\mathbf{V}_j^{\text{opt}}\|_F$.
Return $\mathbf{V}_{j,\text{BB}}$ and $\mathbf{V}_{j,\text{RF}}$.

constraint.

The problem (4.46a)–(4.46c) takes the form of a sparse signal recovery problem [116] and can be efficiently solved using the OMP methods [117]. The steps to obtain the optimal energy-efficient hybrid precoders are outlined in Algorithm 4.2. Note that steps 4–7 find the basis vectors of $\mathbf{A}_{j,\mathrm{T}}$ that have the best representation of $\mathbf{V}_j^{\mathrm{opt}}$, and the transmit power constraint is ensured in step 11 as $\mathbf{V}_j^{\mathrm{opt}}$ has already been designed in section 4.3 to satisfy the power constraint.

4.4.2 Design of hybrid combiners

The EE hybrid combiners at the receive device j are designed by the decomposition of the optimal MMSE receiver $\mathbf{W}_j^{\mathrm{opt}}$ from equation (4.33) into the product of $\mathbf{W}_{j,\mathrm{RF}}$ and $\mathbf{W}_{j,\mathrm{BB}}$. This approach is based on minimising the MMSE between the transmitted and received signal. Hence, a minimisation problem to obtain the hybrid combiners can be formulated as

$$\min_{\mathbf{W}_{\mathrm{RF}},\mathbf{W}_{\mathrm{BB}}} \mathbb{E}\{\|\mathbf{s}_j - \mathbf{W}_{j,\mathrm{BB}}^H \mathbf{W}_{j,\mathrm{RF}}^H \mathbf{y}_j\|^2\} \tag{4.47a}$$

$$\text{s.t.} \qquad \mathbf{W}_{j,\mathrm{RF}} \in \mathcal{W}_{j,\mathrm{RF}}. \tag{4.47b}$$

If we assume $\mathbf{W}_{j,\mathrm{RF}}$ is given, the problem (4.47a)–(4.47b) can be restated as

$$\min_{\mathbf{W}_{\mathrm{BB}}} \mathbb{E}\{\|\mathbf{s}_j - \mathbf{W}_{j,\mathrm{BB}}^H \mathbf{W}_{j,\mathrm{RF}}^H \mathbf{y}_j\|^2\}, \tag{4.48}$$

which has the well-known least-squares solution

$$\mathbf{W}_{j,\mathrm{BB}} = \mathbf{W}_{j,\mathrm{RF}}^H \, \mathbb{E}\{\mathbf{y}_j \mathbf{s}_j^H\} \left(\mathbf{W}_{j,\mathrm{RF}}^H \, \mathbb{E}\{\mathbf{y}_j \mathbf{y}_j^H\} \, \mathbf{W}_{j,\mathrm{RF}}\right)^{-1}. \tag{4.49}$$

The problem (4.47a)–(4.47b) can be rewritten as

$$\min_{\mathbf{W}_{\mathrm{RF}},\mathbf{W}_{\mathrm{BB}}} \|\mathbf{R}_{\mathbf{y}_j,\mathbf{y}_j}^{1/2} \mathbf{W}_j^{\mathrm{opt}} - \mathbf{R}_{\mathbf{y}_j,\mathbf{y}_j}^{1/2} \mathbf{W}_{j,\mathrm{RF}} \mathbf{W}_{j,\mathrm{BB}}\|_F^2 \tag{4.50a}$$

$$\text{s.t.} \qquad \mathbf{W}_{j,\mathrm{RF}} \in \mathcal{W}_{j,\mathrm{RF}}, \tag{4.50b}$$

where $\mathbf{W}_j^{\mathrm{opt}}$ is given by equation (4.33), $\mathbf{R}_{\mathbf{y}_j,\mathbf{y}_j} = \mathbb{E}\{\mathbf{y}_j \mathbf{y}_j^H\}$, $\mathbf{R}_{\mathbf{s}_j,\mathbf{s}_j} = \mathbb{E}\{\mathbf{s}_j \mathbf{s}_j^H\}$, and equation (4.50a) is derived by introducing $\mathrm{tr}\{\left(\mathbf{W}_j^{\mathrm{opt}}\right)^H \mathbf{R}_{\mathbf{y}_j,\mathbf{y}_j} \mathbf{W}_j^{\mathrm{opt}}\} - \mathrm{tr}\{\mathbf{R}_{\mathbf{s}_j,\mathbf{s}_j}\}$ into the objective function (4.47a). Note that since the terms $\left(\mathbf{W}_j^{\mathrm{opt}}\right)^H \mathbf{R}_{\mathbf{y}_j,\mathbf{y}_j} \mathbf{W}_j^{\mathrm{opt}}$ and $\mathbf{R}_{\mathbf{s}_j,\mathbf{s}_j}$ are independent of $\mathbf{W}_{\mathrm{RF}}$ and $\mathbf{W}_{\mathrm{BB}}$, the problems (4.47a)–(4.47b) and

(4.50a)–(4.50b) share the same optimal solutions.

By following steps similar to the design of the hybrid precoders in section 4.4.1, we can remove the non-convexity of the constraint (4.50b) by choosing the array response vectors $\mathbf{a}_{\mathrm{R}}(\phi_{j,\mathrm{R}}^{l,k}, \theta_{j,\mathrm{R}}^{l,k})\ \forall\, l,k$ as the feasible set $\mathcal{W}_{j,\mathrm{RF}}$. Consequently, the equivalent problem is stated as

$$\min_{\mathbf{W}_{\mathrm{BB}}} \|\mathbf{R}_{\mathbf{y}_j,\mathbf{y}_j}^{1/2}\mathbf{W}_j^{\mathrm{opt}} - \mathbf{R}_{\mathbf{y}_j,\mathbf{y}_j}^{1/2}\mathbf{A}_{j,\mathrm{R}}\mathbf{W}_{j,\mathrm{BB}}\|_F^2 \tag{4.51a}$$

$$\text{s.t.}\, \|\mathrm{diag}\left(\mathbf{W}_{j,\mathrm{BB}}\mathbf{W}_{j,\mathrm{BB}}^H\right)\|_0 \leq N_{\mathrm{R}}^{\mathrm{RF}}, \tag{4.51b}$$

where $\mathbf{A}_{j,\mathrm{R}} = [\mathbf{a}_{\mathrm{R}}(\phi_{j,\mathrm{R}}^{1,1}, \theta_{j,\mathrm{R}}^{1,1}), \ldots, \mathbf{a}_{\mathrm{R}}(\phi_{j,\mathrm{R}}^{N_{\mathrm{cl}},N_{\mathrm{p}}}, \theta_{j,\mathrm{R}}^{N_{\mathrm{cl}},N_{\mathrm{p}}})]$, and can be solved by the OMP method, summarized in Algorithm 4.3. For each iteration, steps 4–7 selects the column of $\mathbf{R}_{\mathbf{y}_j,\mathbf{y}_j}^{1/2}\mathbf{A}_{j,\mathrm{R}}$ most correlated to $\mathbf{R}_{\mathbf{y}_j,\mathbf{y}_j}^{1/2}\mathbf{W}_j^{\mathrm{opt}}$ and appends it to $\mathbf{W}_{j,\mathrm{RF}}$. $\mathbf{W}_{j,\mathrm{BB}}$ is derived from the least-squares solution given in equation (4.49). After $N_{\mathrm{R}}^{\mathrm{RF}}$ iterations, the optimal combiners $\mathbf{W}_{j,\mathrm{RF}}$ and $\mathbf{W}_{j,\mathrm{BB}}$ are found and problem (4.47a)–(4.47b) is solved.

Algorithm 4.3: OMP-based algorithm for energy-efficient hybrid combiners.

1 Initialise $\mathbf{W}_{j,\mathrm{RF}}$ and $\mathbf{W}_{j,\mathrm{BB}}$.
2 Set $\mathbf{W}_{j,\mathrm{res}} = \mathbf{W}_j^{\mathrm{opt}}$ using (4.33).
3 **for** $i \leq N_{\mathrm{R}}^{\mathrm{RF}}$ **do**
4 $\quad \mathbf{\Omega}_j = \left(\mathbf{R}_{\mathbf{y}_j,\mathbf{y}_j}^{1/2}\mathbf{A}_{j,\mathrm{R}}\right)^H \left(\mathbf{R}_{\mathbf{y}_j,\mathbf{y}_j}^{1/2}\mathbf{W}_{j,\mathrm{res}}\right)$
5 $\quad [\mathrm{value}, \mathrm{index}] = \max\left(\mathrm{diag}\left(\mathbf{\Omega}_j\mathbf{\Omega}_j^H\right)\right)$
6 $\quad k = \mathrm{index}$
7 $\quad \mathbf{W}_{j,\mathrm{RF}} = \left[\mathbf{W}_{j,\mathrm{RF}}|\mathbf{A}_{j,\mathrm{R}}^{\mathrm{k}}\right]$
8 $\quad$ Update $\mathbf{W}_{j,\mathrm{BB}}$ using (4.49)
9 $\quad$ Update $\mathbf{W}_{j,\mathrm{res}} = \frac{\mathbf{W}_j^{\mathrm{opt}} - \mathbf{W}_{j,\mathrm{RF}}\mathbf{W}_{j,\mathrm{BB}}}{\|\mathbf{W}_j^{\mathrm{opt}} - \mathbf{W}_{j,\mathrm{RF}}\mathbf{W}_{j,\mathrm{BB}}\|_F}$
10 **end**
11 Return $\mathbf{W}_{j,\mathrm{BB}}$ and $\mathbf{W}_{j,\mathrm{RF}}$.

4.4.3 Computational complexity

The following number of floating point operations are required in the evaluation of Algorithm 4.2 to design the proposed hybrid precoders:

- the multiplication of $\mathbf{A}_{j,\mathrm{T}}^H\mathbf{V}_{j,\mathrm{res}}$ in step 4 requires $2N_{\mathrm{T}}^{\mathrm{RF}}N_{\mathrm{T}}N_{\mathrm{s}} - N_{\mathrm{T}}^{\mathrm{RF}}N_{\mathrm{s}}$ flops,

- $2N_{\mathrm{T}}^{\mathrm{RF}}N_{\mathrm{s}}^2 + 2N_{\mathrm{T}}^{\mathrm{RF}}N_{\mathrm{s}} - N_{\mathrm{s}}^2 - Ns$ flops are required to compute diag $\left(\mathbf{\Omega}_j\mathbf{\Omega}_j^H\right)$,
- computing the baseband precoder in step 8 requires $\left(N_{\mathrm{T}}^{\mathrm{RF}}\right)^3 + \left(N_{\mathrm{T}}^{\mathrm{RF}}\right)^2 + N_{\mathrm{T}}^{\mathrm{RF}} + N_{\mathrm{T}}\left(2N_{\mathrm{T}}^{\mathrm{RF}}N_{\mathrm{T}} + 2N_{\mathrm{T}}^{\mathrm{RF}} - N_{\mathrm{T}} - 1\right) + 2N_{\mathrm{T}}^{\mathrm{RF}}N_{\mathrm{s}}\left(N_{\mathrm{T}} + N_{\mathrm{T}}^{\mathrm{RF}} - 1\right)$ flops, and
- updating $\mathbf{V}_{j,\mathrm{res}}$ in step 9 requires $N_{\mathrm{T}}N_{\mathrm{s}}$ flops.

  Therefore, the total number of flops required for computing the precoders $\mathbf{V}_{j,\mathrm{RF}}$ and $\mathbf{V}_{j,\mathrm{BB}}$ is $\mathcal{O}\left(K\left(\left(N_{\mathrm{T}}^{\mathrm{RF}}\right)^3 + 2\left(N_{\mathrm{T}}^{\mathrm{RF}}\right)^2 N_{\mathrm{s}} + 2N_{\mathrm{T}}^{\mathrm{RF}}N_{\mathrm{T}}^2 + 2N_{\mathrm{T}}^{\mathrm{RF}}N_{\mathrm{s}}^2\right)\right)$.

The computational complexity of the hybrid combiners in Algorithm 4.3 can be obtained in a similar manner.

4.5 Imperfect CSI Considerations

In the previous sections, the near optimal EE maximisation based beamformers were designed under the assumption of perfect CSI at the IoT nodes. In practice, however, it may not always be possible to obtain perfect CSI at all the IoT nodes due to errors in channel estimation, quantisation errors and feedback delay, amongst other factors. Therefore, in this section we investigate to what extent the analysis can be extended for the imperfect CSI case. In particular, we consider the design of robust hybrid beamformers which account for the fact that the available channel knowledge would be based on estimates obtained from an imperfect observation of the actual channel state.

In this regard, the imperfect channel is modelled as

$$\mathbf{H}_{j,j} = \hat{\mathbf{H}}_{j,j} + \Delta_j, \tag{4.52}$$

$$\mathbf{H}_{k,j} = \hat{\mathbf{H}}_{k,j} + \Delta_k \;\; k \neq j, \tag{4.53}$$

where $\mathbf{H}$ and $\hat{\mathbf{H}}$ denote the perfect and imperfect channels, respectively, and Δ represents the CSI stochastic error, which is assumed to be independent of the perfect channel and is distributed as

$$\Delta \sim \mathcal{CN}\left(0, \gamma\mathbf{I}\right), \tag{4.54}$$

where γ represents the variance of the CSI error given by [118]

$$\gamma \triangleq \tau\rho^{-\nu} \;\; \tau > 0, \nu \geq 0, \tag{4.55}$$

with the parameters τ and ν denoting various instances of CSI acquisition and ρ representing the signal-to-noise-ratio of the link given as

$$\rho = P_j \|\mathbf{\Sigma}_j^{-1}\mathbf{H}_{j,j}\mathbf{V}_j\mathbf{V}_j^H\mathbf{H}_{j,j}^H\|_F, \tag{4.56}$$

where $\mathbf{\Sigma}_j$ is expressed as

$$\begin{aligned}\mathbf{\Sigma}_j &= \kappa\mathbf{H}_{j,j}\text{diag}\left(\mathbf{V}_j\mathbf{V}_j^H\right)\mathbf{H}_{j,j}^H + \sum_{k\neq j}^{K}\mathbf{H}_{k,j}\left(\mathbf{V}_k\mathbf{V}_k^H + \kappa\,\text{diag}\left(\mathbf{V}_k\mathbf{V}_k^H\right)\right)\mathbf{H}_{k,j}^H \\ &+ \beta\,\text{diag}\left(\mathbf{H}_{j,j}\mathbf{V}_j\mathbf{V}_j^H\mathbf{H}_{j,j}^H\right) + \sum_{k\neq j}^{K}\beta\,\left(\text{diag}\left(\mathbf{H}_{k,j}\mathbf{V}_k\mathbf{V}_k^H\mathbf{H}_{k,j}^H\right)\right) + \sigma_z^2\mathbf{I}_{N_\text{R}},\end{aligned} \tag{4.57}$$

indicating that the stochastic error is affected by the hardware impairments. Since all the nodes have access to $\hat{\mathbf{H}}$ instead of $\mathbf{H}$, rather than focusing on the actual achievable EE we consider the lower bound, EE^L, where the estimation errors are treated as noise [119]. Accordingly,

$$\begin{aligned}\text{EE}^\text{L}(\mathbf{V}) &= \frac{f_1(\mathbf{V})}{f_2(\mathbf{V})} \\ &= \frac{\sum_{j=1}^{K}\varphi_j\mathcal{I}_j^\text{L}(\mathbf{V})}{\sum_{j=1}^{K}\left[\vartheta\text{tr}\{\mathbf{V}_j\mathbf{V}_j^H\} + N_\text{T}^{j,\text{RF}}P_\text{RF} + N_{j,\text{PS}}P_\text{PS} + P_\text{sta}\right]},\end{aligned} \tag{4.58}$$

where $\mathcal{I}_j^\text{L} = \log\left|\mathbf{I}_{N_\text{R}} + \hat{\mathbf{\Sigma}}_j^{-1}\hat{\mathbf{H}}_{j,j}\mathbf{V}_j\mathbf{V}_j^H\hat{\mathbf{H}}_{j,j}^H\right|$ is the rate and $\hat{\mathbf{\Sigma}}_j$ is defined as

$$\begin{aligned}\hat{\mathbf{\Sigma}}_j &= \kappa\hat{\mathbf{H}}_{j,j}\text{diag}\left(\mathbf{V}_j\mathbf{V}_j^H\right)\hat{\mathbf{H}}_{j,j}^H + \sum\nolimits_{k\neq j}^{K}\hat{\mathbf{H}}_{k,j}\left(\mathbf{V}_k\mathbf{V}_k^H + \kappa\,\text{diag}\left(\mathbf{V}_k\mathbf{V}_k^H\right)\right)\hat{\mathbf{H}}_{k,j}^H \\ &+ \beta\,\text{diag}\left(\hat{\mathbf{H}}_{j,j}\mathbf{V}_j\mathbf{V}_j^H\hat{\mathbf{H}}_{j,j}^H\right) + \sum\nolimits_{k\neq j}^{K}\beta\,\left(\text{diag}\left(\hat{\mathbf{H}}_{k,j}\mathbf{V}_k\mathbf{V}_k^H\hat{\mathbf{H}}_{k,j}^H\right)\right) \\ &+ \left(\sigma_z^2 + f_e\right)\mathbf{I}_{N_\text{R}},\end{aligned} \tag{4.59}$$

where f_e reflects the effect of the imperfect CSI and is given by

$$f_e \approx \gamma\left(1+\kappa+\beta\right)\sum\nolimits_{j=1}^{K}\text{diag}\left(\mathbf{V}_j\mathbf{V}_j^H\right). \tag{4.60}$$

Next, the EE maximization problem we want to solve is expressed as

$$\max_{\mathbf{V}} \ \mathrm{EE}^{L}(\mathbf{V}) \tag{4.61a}$$

$$\text{s.t. } \mathrm{tr}\{\mathbf{V}_j\mathbf{V}_j^H\} \leq P_j, \ \forall\, j. \tag{4.61b}$$

Similar to the perfect CSI case, we exploit the relationship between the fractional and parametric programming problems to reformulate the problem as

$$\max_{\mathbf{V}} \ \sum\nolimits_{j=1}^{K} \varphi_j \mathcal{I}_j^{\mathrm{L}}(\mathbf{V}) - g f_2(\mathbf{V}) \tag{4.62a}$$

$$\text{s.t. } \mathrm{tr}\{\mathbf{V}_j\mathbf{V}_j^H\} \leq P_j, \ \forall\, j, \tag{4.62b}$$

then we will solve this problem by transforming the WSR in equation (4.62a) into an equivalent WMMSE one. This relationship is expressed by the following theorem.

Theorem 4.2. *The imperfect CSI error WSR problem in equations* (4.62a)–(4.62b) *is equivalent to the WMMSE problem described in equations* (4.63a)–(4.63b) *since the optimal solution for the precoders obtained from solving both problems are identical.*

$$\max_{\mathbf{V},\mathbf{W},\mathbf{Q}} \ \sum\nolimits_{j=1}^{K} \left[tr\{\mathbf{Q}_j^{L}\mathbf{M}_j^{\mathrm{L}}\} - \varphi_j \log\left|\frac{\ln 2}{\varphi_j}\mathbf{Q}_j^{L}\right| - \frac{N_{\mathrm{s}}\varphi_j}{\ln 2}\right] - g f_2(\mathbf{V}) \tag{4.63a}$$

$$s.t. \quad tr\{\mathbf{V}_j\mathbf{V}_j^H\} \leq P_j, \ \forall\, j. \tag{4.63b}$$

Proof. Using $\mathbf{Q}_j^{\mathrm{L}}$ as defined in equation (4.64) and following steps analogous to Theorem. 4.1, the proof can be obtained.

$$\mathbf{Q}_j^{\mathrm{L}} = \frac{\varphi_j}{\ln 2}\left(\mathbf{E}_j^{\mathrm{L}}(\mathbf{V})\right)^{-1}, \tag{4.64}$$

where $\mathbf{E}_j^{\mathrm{L}}$ is defined as

$$\mathbf{E}_j^{\mathrm{L}} = \left(\mathbf{I}_{N_{\mathrm{s}}} + \mathbf{V}_j^H\hat{\mathbf{H}}_{j,j}^H\hat{\mathbf{\Sigma}}_j^{-1}\hat{\mathbf{H}}_{j,j}\mathbf{V}_j\right). \tag{4.65}$$

Note that for the imperfect CSI problem, the optimal receiver at device j is given by

$$\mathbf{W}_j^{\mathrm{L}} = \mathbf{V}_j^H\hat{\mathbf{H}}_{j,j}^H\left(\hat{\mathbf{H}}_{j,j}\mathbf{V}_j\mathbf{V}_j^H\hat{\mathbf{H}}_{j,j}^H + \mathbf{G}_j + \left(\sigma_z^2 + f_e\right)\mathbf{I}_{N_r}\right)^{-1}, \tag{4.66}$$

where $\mathbf{G}_j$ is given by

$$\begin{aligned}\mathbf{G}_j &= \kappa\hat{\mathbf{H}}_{j,j}\mathrm{diag}\left(\mathbf{V}_j\mathbf{V}_j^H\right)\hat{\mathbf{H}}_{j,j}^H + \sum\nolimits_{k\neq j}^{K}\hat{\mathbf{H}}_{k,j}\left(\mathbf{V}_k\mathbf{V}_k^H + \kappa\,\mathrm{diag}\left(\mathbf{V}_k\mathbf{V}_k^H\right)\right)\hat{\mathbf{H}}_{k,j}^H \\ &+ \beta\,\mathrm{diag}\left(\hat{\mathbf{H}}_{j,j}\mathbf{V}_j\mathbf{V}_j^H\hat{\mathbf{H}}_{j,j}^H\right) + \sum\nolimits_{k\neq j}^{K}\beta\,\left(\mathrm{diag}\left(\hat{\mathbf{H}}_{k,j}\mathbf{V}_k\mathbf{V}_k^H\hat{\mathbf{H}}_{k,j}^H\right)\right) \qquad (4.67)\end{aligned}$$

Comparable to the perfect case, we employ an alternating optimization method to solve the stochastic CSI error problem described by equations (4.63a)–(4.63b). Therefore, the optimal precoder at device j can be obtained for fixed $\mathbf{Q}_j^{\mathrm{L}}$ and $\mathbf{W}_j^{\mathrm{L}}$ using the Lagrangian approach. Therefore, we obtain

$$\mathbf{V}_j^{\mathrm{L}} = \left(\lambda_j^{\mathrm{L}}\mathbf{I}_{N_{\mathrm{T}}} + g\vartheta\mathbf{I}_{N_{\mathrm{T}}} + \mathbf{\Phi}_j^{\mathrm{L}}\right)^{-1}\hat{\mathbf{H}}_{j,j}^H\mathbf{W}_j^H\mathbf{Q}_j, \qquad (4.68)$$

where λ_j^{L} is the Lagrangian multiplier associated with the stochastic CSI error problem and $\mathbf{\Phi}_j^{\mathrm{L}}$ is defined as

$$\begin{aligned}\mathbf{\Phi}_j^{\mathrm{L}} &= \hat{\mathbf{H}}_{j,j}^H\mathbf{W}_j^H\mathbf{Q}_j\mathbf{W}_j\hat{\mathbf{H}}_{j,j} + \kappa\,\mathrm{diag}\left(\hat{\mathbf{H}}_{j,j}^H\mathbf{W}_j^H\mathbf{Q}_j\mathbf{W}_j\hat{\mathbf{H}}_{j,j}\right) \qquad (4.69) \\ &+ \beta\hat{\mathbf{H}}_{j,j}^H\mathrm{diag}\left(\mathbf{W}_j^H\mathbf{Q}_j\mathbf{W}_j\right)\hat{\mathbf{H}}_{j,j} + \sum\nolimits_{j=1}^{K}\gamma\left(1+\kappa+\beta\right)\mathrm{diag}\left(\mathbf{W}_j^H\mathbf{Q}_j\mathbf{W}_j\right).\end{aligned}$$

Next, having obtained the optimal precoder $\mathbf{V}_j^{\mathrm{L}}$, the hybrid precoders and combiners of the imperfect CSI error model, $\mathbf{V}_{j,\mathrm{RF}}^{\mathrm{L}}$ and $\mathbf{V}_{j,\mathrm{BB}}^{\mathrm{L}}$, can be derived by following similar steps to algorithms 4.2 and 4.3 of the perfect CSI case. In addition, the convergence and complexity considerations of the perfect CSI formulation are also applicable to the alternating optimization method applied to solve the imperfect CSI error EE maximization problem in equations (4.62a)–(4.62b).

4.6 Numerical Results

In this section, we provide numerical results to show the performance of the proposed energy-efficient hybrid precoders and combiners, and we investigate the EE for a mmWave MIMO IoT network as a function of the hardware impairment parameters (κ, β) and the transmit power constraint for each device i.e., $P_j\ \forall\, j$. The results validate the analyses presented in sections 4.3 and 4.4.

Without loss of the generality, the parameters used in the simulations are as follows. We set the same number of receive and transmit RF chains at each

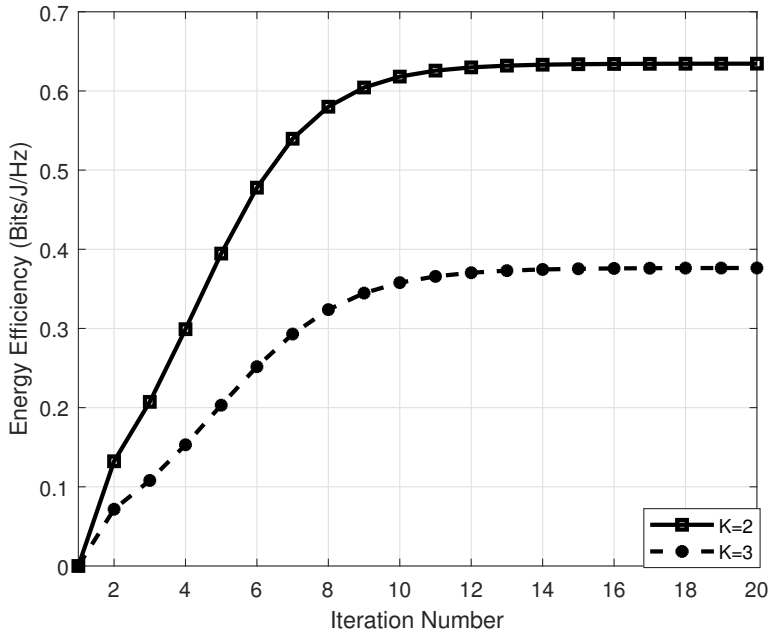


Figure 4.3: Convergence behaviour of Algorithm 4.1 for varying K IoT devices.

device i.e., $N_{\mathrm{R}}^{\mathrm{RF}} = N_{\mathrm{T}}^{\mathrm{RF}} = N^{\mathrm{RF}}$. We also assume an equal number of transmit and receive antennas, i.e., $N_{\mathrm{T}} = N_{\mathrm{R}} = N$, and equal weights for each receive device, i.e., $\varphi_j = \varphi \ \forall\, j$. The initial values for the variables in the proposed algorithms are randomly generated, convergence thresholds are set to 10^{-5}, noise variance σ_z^2 is fixed at 1, and all the results presented are averaged over 100 channel realizations. Unless otherwise stated, some of the simulation parameters are: $\kappa = \beta = -40$ dB indicating a wide dynamic range [120, 121], $N_{\mathrm{RF}} = 2$, $N = 4$, $\vartheta = 1/0.32$, $N_{\mathrm{cl}} = 8$, $N_{\mathrm{p}} = 10$, $N_{\mathrm{s}} = K$ and $P_j = 40$ dBm.

We begin by illustrating the evolution of the WMMSE-based EE maximisation algorithm described by Algorithm 4.1 in Fig. 4.3. We set the dynamic circuit power to $P_{\mathrm{RF}} = 20$ dBm and $P_{\mathrm{PS}} = 0$ dBm. The consumed static power is set to $P_{\mathrm{sta}} = 27$ dBm. It can be seen from the figure that the alternating maximisation process converges in 10–12 iterations when $K = 2$, whereas when $K = 3$ the process converges in 13–15 iterations. Nevertheless, for both cases the monotonic increase of the EE is verified.

Next, in Fig. 4.4, we analyse the effect of the transmit and receive distortions on the mmWave IoT network. Accordingly, the EE is plotted with respect to $\kappa = \beta$ values. The settings for the dynamic and static power consumption of the transmitting devices are maintained from Fig. 4.3. Here, we illustrate that high

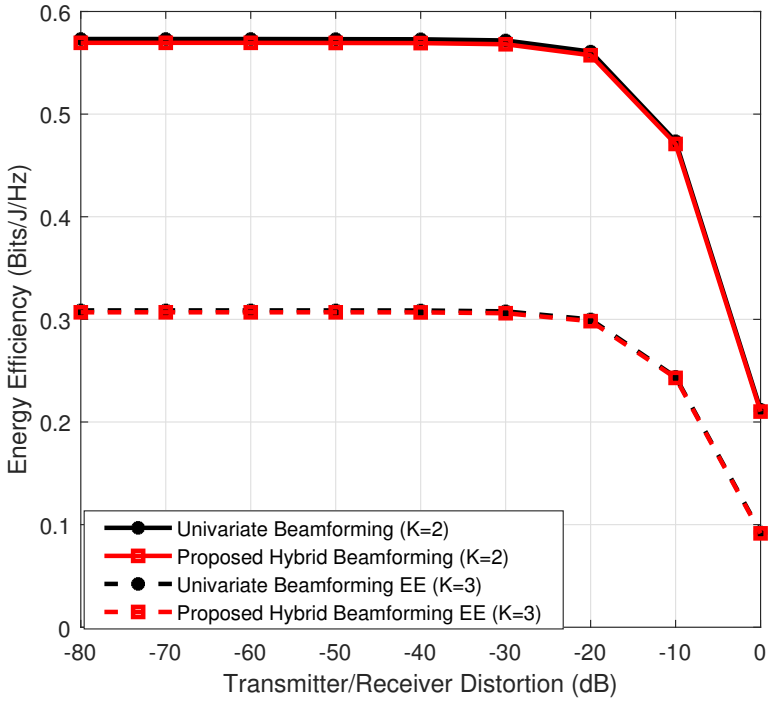


Figure 4.4: Comparison of EE achieved by transceiver design solutions with respect to transmitter/receiver distortions, i.e., $\kappa = \beta$ for $K = 2$ and $K = 3$.

distortions at the transmitter/receiver lead to a decrease in EE since the mmWave system requires more power to achieve the transmission requirement. Moreover, it can also be seen that our proposed hybrid transceiver design algorithm achieves virtually the same EE as the optimal univariate design. To quantify this, the hybrid transceiver design algorithm achieves up to 99 % of the optimal univariate performance. This result validates Algorithms 4.2 and 4.3. Furthermore, as can be seen in Fig. 4.4, the EE maintained at low transmitter/receiver distortions is about 0.58 bits/J/Hz for $K = 2$, which is slightly higher than the EE of 0.31 bits/J/Hz when $K = 3$. This outcome is due to the fact that the interference in the system increases with an increase in the number of devices, which directly affects the power allocation in the beamforming solution.

In Fig. 4.5, we compare the EE performance achieved by employing the hybrid transceiver design for two different designs when $K = 2$, namely (1) SE maximisation and (2) EE maximisation. The SE maximisation can be achieved by a slight modification to Algorithm 4.2. Specifically, when the EE parameter g is not updated, steps 1–6 solve the WSR problem. Hence, Fig. 4.5 illustrates the SE and EE performances with respect to maximum transmit power. It can

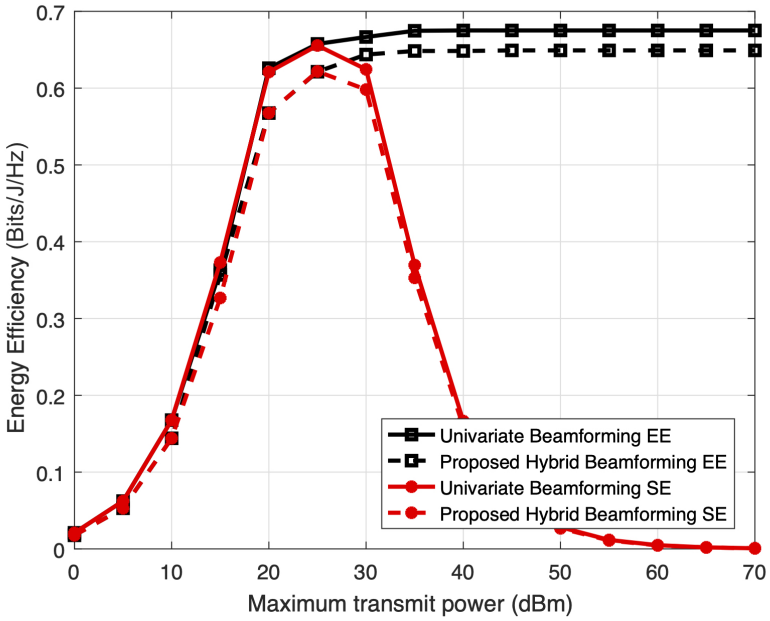


Figure 4.5: Comparison of EE achieved by EE maximisation and SE maximisation design criteria for $K = 2$.

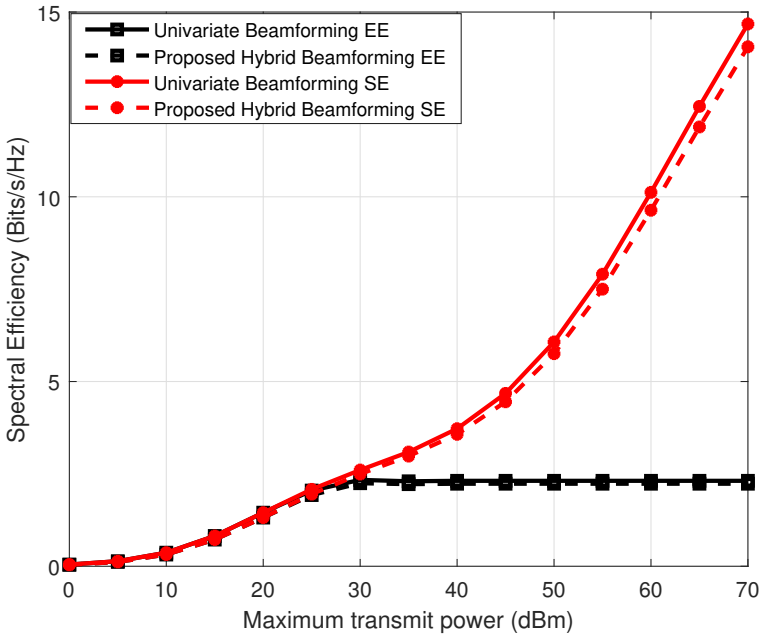


Figure 4.6: Comparison of SE achieved by EE maximisation and SE maximisation design criteria for $K = 2$.

be observed that in the low transmit power regime, both designs result in equivalent EE performance, i.e., an increase in the maximum transmit power yields a

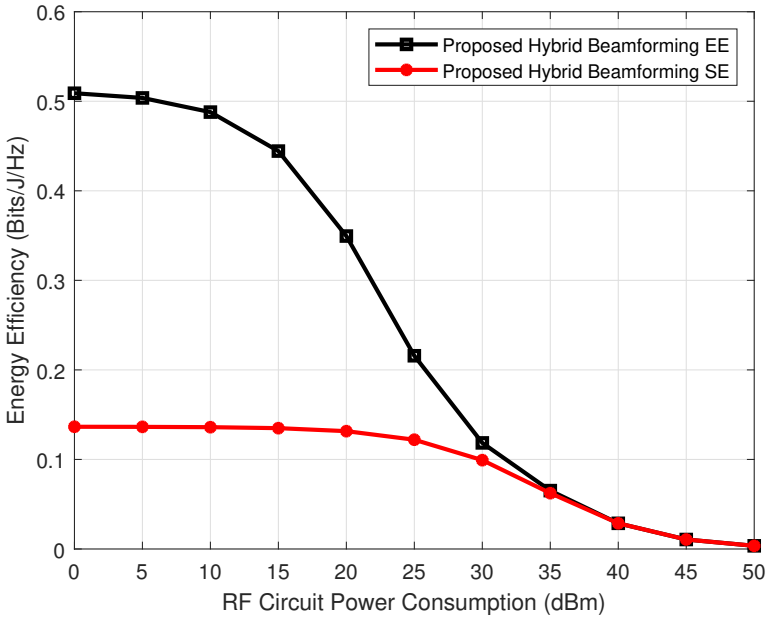


Figure 4.7: Comparison of EE achieved by EE and SE maximisation design solutions with respect to circuit power consumption

corresponding rise in EE. However, the performances of the two designs diverge drastically as the transmit power is increased. In particular, the EE resulting from the SE maximisation design steadily decreases after achieving a maximum value of approximately 0.7 bits/J/Hz at 25 dBm. This can be explained by the fact that maximising the SE requires full transmit power, which then leads to an increase in total power consumption. Conversely, the EE obtained from the EE maximisation design remains constant beyond the peak value obtained at 25 dBm, as this design takes into account the total power consumption of the system. In other words, the EE maximisation design guarantees that the optimal transmit power is achieved. Additionally, the small gap in performance between the results obtained by the optimal univariate transceiver design and that produced by the proposed algorithms validates the hybrid transceiver design.

Next, for the sake of comparison, we evaluate the SE performance of the system with respect to the maximum transmit power constraint in Fig. 4.6 for both the EE and SE maximisation designs. A monotonic increase in the SE performance is achieved by the SE maximisation design, in direct proportion to the maximum transmit power. This result is expected since the SE maximisation design fails to take into consideration the total power consumption of the system and is just focussed on increasing the SE of the system. Thus, although a high

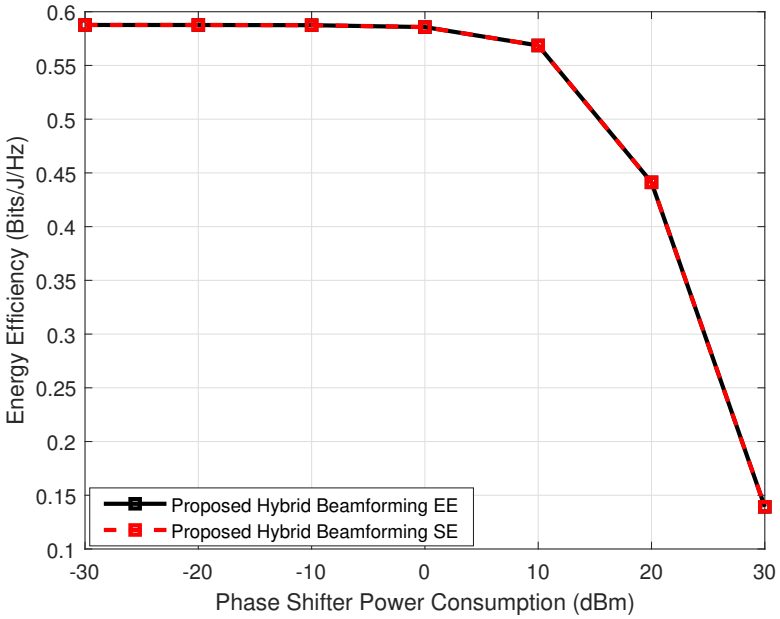
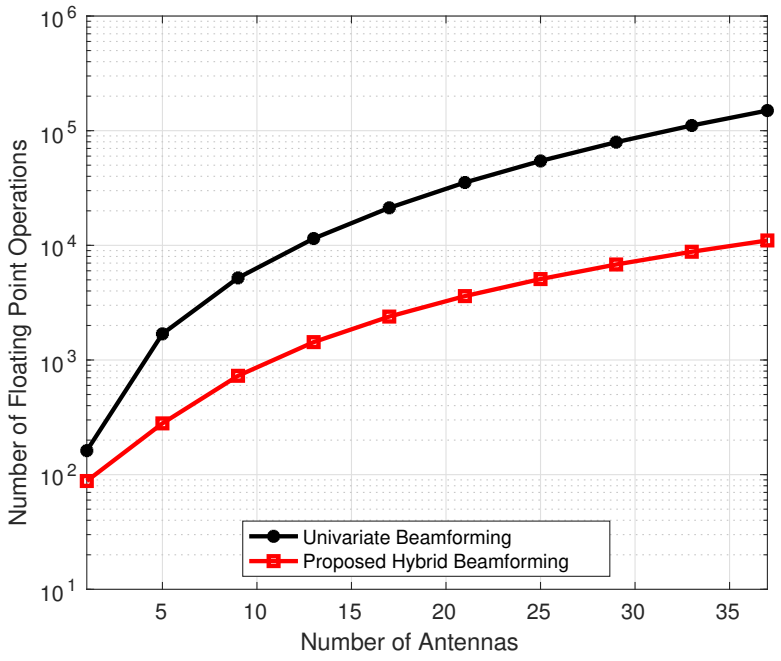


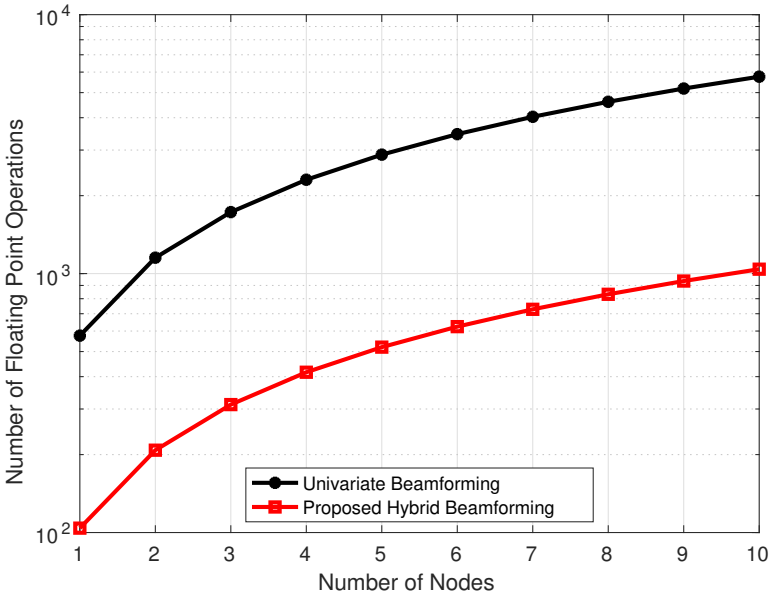
Figure 4.8: EE achieved by the EE maximisation hybrid transceiver design with respect to phase shifter power consumption.

transmit power implies high circuit power consumption, this is not accounted for in this design. In contrast, the SE obtained by the EE maximisation design increases until the optimal transmit power is achieved, then it remains constant irrespective of further increase in the transmit power. Specifically, after 25 dBm, which is the maximum transmit power for the system, the spectral efficiency saturates at around 2 bits/s/Hz. Hence, we illustrate by this result that if the maximum transmit power is higher than the optimal transmit power for a given system, the EE maximisation design ensures that transmission happens at or below the optimal power so as to provide the maximum EE. This result highlights the benefits of the proposed optimal EE maximisation design in mmWave IoT networks.

In Fig. 4.7, we evaluate the EE performance of both designs with respect to the power consumed by the RF components of the system (i.e., RF chains). We set the dynamic power consumed by the transmitters to $P_{\text{RF}} = 20$ dBm and $P_{\text{PS}} = 0$ dBm, the static power to $P_{\text{sta}} = 27$ dBm, the number of devices $K = 2$, and the maximum transmit power, $P_j = 40$ dBm. It can be seen that the hybrid transceiver design based on EE maximisation produces better EE performance than the results from the SE maximisation design, especially for low circuit power consumption. However, the EE performance of both the designs at high circuit



(a)



(b)

Figure 4.9: Complexity comparisons of univariate and proposed hybrid transceiver designs with respect to (a) number of antennas and (b) number of devices.

power consumption becomes identical. This is due to the fact that when the total power consumed is mostly the circuit power, maximising the EE becomes equivalent to the maximisation of SE.

In Fig. 4.8, we analyse the impact of the power consumption of the phase shifters on the EE of the mmWave IoT system. The settings of the dynamic and static power consumption, and maximum transmit power are maintained from Fig. 4.7. Here, we observe that an EE of around 0.6 bits/J/Hz is sustained for low phase shifter power consumption. However, the EE decreases with an increase in the phase shifter power consumption. This result suggests that low power phase shifters are required in practical mmWave IoT systems for energy-efficient operation. In addition, the identical EE performance of both transceiver designs suggests that the contribution of the phase shifters to the total power consumed when compared with the contribution from the other sources is minimal.

Now that we have established the fact that the proposed hybrid transceiver design algorithms produce near-optimal performance, we present the computational complexity comparison of the proposed hybrid algorithms (Algorithms 4.2 and 4.3) with the WMMSE-based univariate transceiver design algorithm (Algorithm 4.1) for different numbers of antennas and devices in Fig. 4.9(a) and Fig. 4.9(b) respectively. We observe from both figures that the computational complexity of the proposed hybrid transceiver design algorithm is much lower than that of the univariate algorithm, especially when the number of antennas and devices is high. This result further illustrates the benefits of the proposed EE maximisation design for IoT networks.

Finally, to show how the robust EE design performs, Fig. 4.10 illustrates the energy and spectral efficiencies for varying values of τ from the imperfect CSI error model discussed in section 4.5. We set the dynamic power consumed by the transmitters to $P_{\mathrm{RF}} = 20$ dBm and $P_{\mathrm{PS}} = 0$ dBm, the static power to $P_{\mathrm{sta}} = 27$ dBm, the number of devices $K = 2$, and the maximum transmit power $P_j = 30$ dBm. The robust design is compared with the non-robust version, obtained by using the available imperfect channel as if it were perfect. It can be seen that for a fixed value of ν, both energy and spectral efficiencies decrease with increasing values of τ. This can be explained by the fact that larger τ values imply an increase in the CSI error variance, thus the performance degrades with more errors. Additionally, it can be noticed $\tau = 0$ corresponds to the perfect CSI case, as shown by the equivalent performances of the non-robust and robust

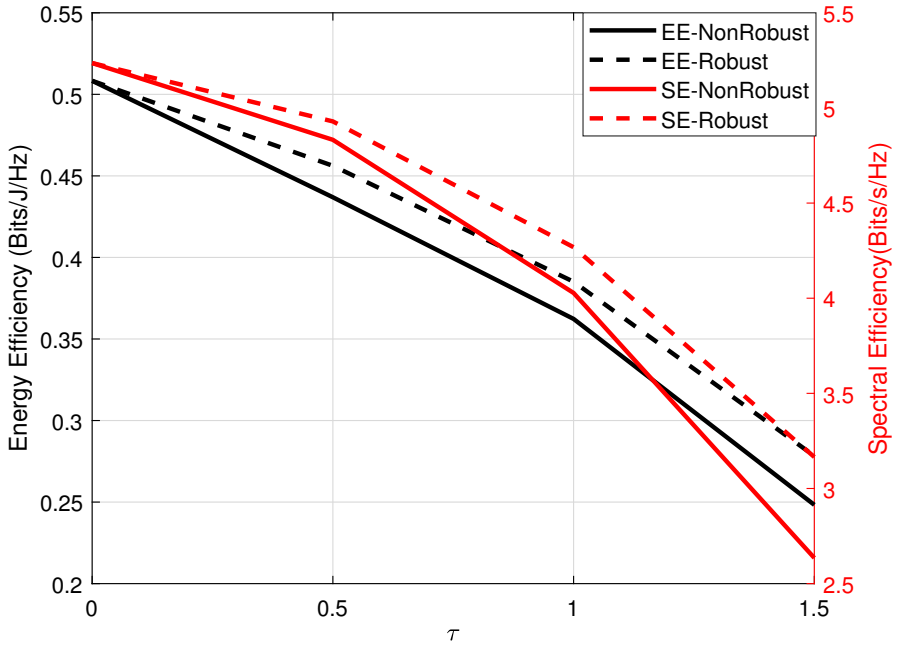


Figure 4.10: Comparison of EE and SE achieved by hybrid transceiver design with respect to varying imperfect CSI errors.

designs. However, as τ increases, the proposed robust design outperforms the non-robust design in both EE and SE, stressing the resilience of the proposed robust beamforming designs for mmWave IoT networks.

Remark 4.2. *Note that EE values ranging from* $0-1$ *are common place in literature, (e.g. [122, 123, 124, 125, 126].), where the EE is computed as a ratio of the spectral efficiency (bits/s/Hz) to the energy consumed (J/s).*

4.7 Conclusion

In this chapter, we studied an energy-efficient hybrid transceiver design for mmWave interference channels involving IoT devices, taking into consideration the inevitable hardware impairments at the transmitters and receivers. Since the joint hybrid transceiver design problems were non-convex, we transformed them into their equivalent univariate maximisation problems and proposed alternating maximisation algorithms based on the relationship between WSR and WMMSE to solve them. Furthermore, using the concept of orthogonal matching pursuit, we proposed a low-complexity optimal hybrid transceiver design algorithm that

maximises the EE of the IoT network. Numerical results showed that the proposed algorithms achieve near-optimal EE performance to support mmWave IoT systems for both perfect and imperfect CSI considerations.

Chapter 5

Rate-Splitting Transmission in Multi-User mmWave Systems

5.1 Introduction

As discussed earlier in Chapter 2, the tremendous interest of recent years in mmWave has resulted in major advances in terms of understanding mmWave cellular communication. While these initial studies have provided a baseline for mmWave propagation, they have mainly been studied under the idealistic assumption of the availability of perfect CSI. However, in practice, the available CSI is usually obtained by imperfect estimates which result in performance loss. Therefore, in this chapter we propose an alternative transmission strategy that will enable us to glean more improvements in the system's performance despite the availability of imperfect CSI.

Within the context of imperfect CSI, some works have focused on estimation of the mmWave channel (the research background for channel estimation was provided earlier in section 2.4). Other works have focused on hybrid beamforming design that avoids the dependence on perfect CSI. For example, [44] proposed the hybrid transceiver design in a single-user scenario that relies on partial channel knowledge at the BS and UE. For settings with multiple users, the hybrid beamforming approach proposed in [80] relies on a two-stage feedback method, where the analog beamformer is designed using instantaneous CSI at the receiver, and the baseband beamformer is designed by the quantized CSI containing in-

Work in this chapter has been presented at IEEE SPAWC, 2018 [127].

formation about an effective channel at the transmitter. In addition, a simplified signalling and feedback procedure was introduced in [128] in which analog beamforming directions are designed by feedback from the receiver and second-order statistics of the channel is used to tackle the MUI at the baseband level. These studies illustrate the usability of mmWave communication for practical scenarios. Nevertheless, if CSI inaccuracies are considered, the residual MUI is a restricting factor to mmWave systems. To avoid this, we investigate the RS transmission strategy in this chapter.

In the RS approach, successive interference cancellation (SIC) and superposition coding techniques are utilised. In particular, for the multi-user system setting, the message intended for each UE is divided into two parts: a common part and a private part. This common message is drawn from a public codebook accessible to all UEs in the system and should therefore be decoded by all the UEs with zero error probability. The private messages are transmitted using a fraction of the total power while the common part is superimposed on the private messages by using the residual power. At the user end, the common message is decoded first by treating the private messages as noise and then removed from the received signal using SIC. Afterwards, each UE decodes its desired private message from the received signal.

RS has been studied extensively for scenarios with imperfect CSI in standard multi-user MIMO communications and is shown to be robust to CSI errors since the attainability of the degree of freedom (DoF) of the common message is not dependent on interference cancellation. Specifically, RS was shown to expand the achievability of the degrees of freedom (DoFs) when CSI inaccuracies increase for multiple-input single-output (MISO) broadcast channels[129, 130], and for MIMO systems [131]. Compared with conventional transmission methods that do not employ RS (these methods will subsequently be referred to as "NoRS"), RS has been shown to provide a reduction in feedback overhead in [132]. Furthermore, for MIMO settings that account for hardware impairments, RS is introduced to mitigate the effects of phase and amplified thermal noises in [133]. Despite the fact that RS has been considerably researched for standard MIMO systems, few studies have attempted employing the RS transmission strategy for mmWave transmissions, as applying RS strategy to multi-user mmWave MIMO systems under imperfect CSI, where the users are equipped with multiple antennas, is non-trivial.

Accordingly, in this work, we build a tractable mmWave system investigating the RS transmissions and characterising the spectral efficiency performance achieved by this strategy. At this point, we would like to note that our work differs from [128], where RS was introduced for multi-user mmWave system in the presence of statistical CSIT, due to the fact that we consider multiple antennas at the UEs. Additionally, the channel estimation technique presented in this chapter differs from the one-stage feedback scheme considered in [128], in that we consider quantised CSI feedback using adaptive compressed sensing tools. The results illustrate that transmission by means of RS provides significant gains in achievable sum rate over the standard NoRS transmission strategy. In addition, it is also shown that RS performance can be improved by increasing the resolution parameter. Nevertheless, RS is robust to the available level of channel knowledge.

The rest of this chapter is organised as follows. Section 5.2 presents the some preliminaries including the system model, non-ideal hardware and imperfect CSI considerations and hybrid beamforming with the RS transmission strategy. Section 5.3 provides the design of the RF, BB and RS precoders and characterises the SINR and rate performances of the mmWave system under imperfect CSI and non-ideal hardware. Simulation results are presented in section 5.4, and finally section 5.5 contains a summary of the chapter.

5.2 System Model

In this section, we illustrate our system model of a multi-user mmWave downlink system where the BS equipped with N_{T} antennas and $N_{\mathrm{T}}^{\mathrm{RF}}$ RF chains transmits simultaneously to K N_{R}-antenna users by means of N_{s} data streams, as illustrated by Fig. 5.1.

The mmWave channels are expected to be sparse [13] and we assume a geometric channel model with a few scatterers, each of which contributes a propagation path between the BS and users. Under this model, the $N_{\mathrm{R}} \times N_{\mathrm{T}}$ channel matrix for user k is expressed as

$$\mathbf{H}_k = \sqrt{\frac{N_{\mathrm{T}} N_{\mathrm{R}}}{W_k}} \sum_{l=1}^{W_k} \alpha_{l,k}\, \mathbf{a}_{\mathrm{UE}}(\phi_{\mathrm{R}}^{l,k}, \theta_{\mathrm{R}}^{l,k}) \mathbf{a}_{\mathrm{BS}}^{H}(\phi_{\mathrm{T}}^{l,k}, \theta_{\mathrm{T}}^{l,k}), \tag{5.1}$$

where W_k is the number of paths between the BS and user k, $\alpha_{l,k} \sim \mathcal{CN}(0,1)$ is the complex gain of the lth path, and the variables $\theta_{\mathrm{R}}^{l,k}$ $(\phi_{\mathrm{R}}^{l,k}), \in [0, 2\pi]$ and

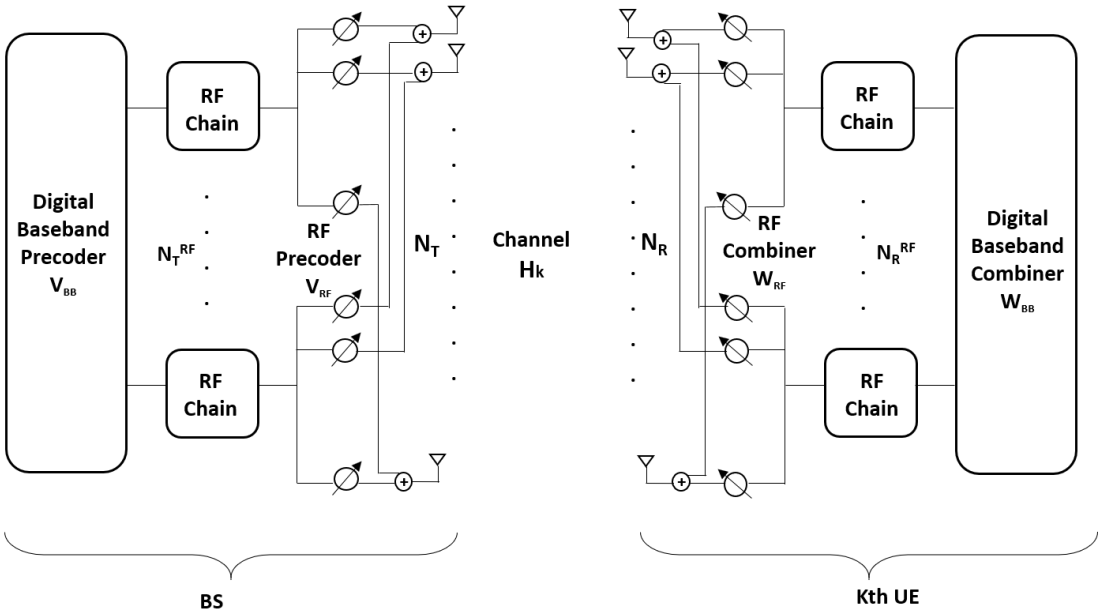


Figure 5.1: Transceiver block diagram of a BS communicating with the kth UE that uses RF and baseband beamformers at both ends.

$\theta_{\mathrm{T}}^{l,k}$ ($\phi_{\mathrm{T}}^{l,k}$) $\in [0, 2\pi]$ correspond to the elevation (azimuth) AoAs and AoDs of the lth path, respectively. Further, $\mathbf{a}_{\mathrm{UE}}$ and $\mathbf{a}_{\mathrm{BS}}$ correspond to the UPA response vectors at the user k and BS, respectively.

5.2.1 Imperfect CSI considerations

As highlighted earlier in Chapter 2, the huge cost of feedback makes obtaining perfect CSI at the BS infeasible. Thus, the BS relies on information from the user to estimate the channel and design the hybrid precoders for transmission. Hence, the geometric channel for user k in equation (5.1) can be estimated by using random measurement matrices to estimate the channel parameters, i.e., the AoAs $(\phi_{\mathrm{R}}^{l,k}, \theta_{\mathrm{R}}^{l,k})$, AoDs $(\phi_{\mathrm{T}}^{l,k}, \theta_{\mathrm{T}}^{l,k})$ and gains $(\alpha_{l,k})$ of each path. Specifically, at the estimation phase, a vector of N_{s} pilot symbols $\mathbf{s}_{\mathrm{p}}$ is transmitted to each user such that $\mathbb{E}\left[\|\mathbf{s}_{\mathrm{p}}\|_2^2\right] = 1$.

Relying on feedback from the user, the hybrid precoders are designed using the multi-resolution codebook structure [60]. More specifically, the RF beamformers are first designed using a codebook beam-training strategy, where through an iterative process of information exchange between the BS and the users, the BS chooses the beam patterns that maximise the power for each user. Next, the effective channel is quantized using random vector quantization and fed back

to the BS for the baseband precoder design. In other words, the BS uses the channel estimates from the user to mitigate MUI. The detailed design of the hybrid precoders are discussed in section 5.3.3.

5.2.2 Conventional hybrid beamforming (NoRS)

We begin by describing the conventional transmission strategy for the mmWave multi-user systems with hybrid beamforming to facilitate our understanding of the similarities and differences between both transmission schemes.

On the downlink, the BS employs a $KN_{\mathrm{T}} \times KN_{\mathrm{T}}^{\mathrm{RF}}$ hybrid RF precoder $\mathbf{V}_{\mathrm{RF}}$, and a $KN_{\mathrm{T}}^{\mathrm{RF}} \times KN_{\mathrm{s}}$ baseband precoder $\mathbf{V}_{\mathrm{BB}}$. Thus, the transmitted signal is given by

$$\mathbf{x} = \mathbf{V}_{\mathrm{RF}}\,\mathbf{V}_{\mathrm{BB}}\,\mathbf{s} = \sum_{k=1}^{K} \sqrt{P_k}\mathbf{V}_{\mathrm{RF}}\mathbf{V}_{\mathrm{BB}}^{k}\mathbf{s}_k, \tag{5.2}$$

where $\mathbf{s} = [\mathbf{s}_1, \ldots, \mathbf{s}_k, \ldots, \mathbf{s}_K]^T$ is the $KN_{\mathrm{s}} \times 1$ vector of transmit data symbols for all users, P_k is the transmit power allocated to user k, and $\mathbf{s}_u$ is the $N_{\mathrm{s}} \times 1$ symbol vector for user u. $\mathbf{V}_{\mathrm{RF}}$ and $\mathbf{V}_{\mathrm{BB}}$ are diagonal matrices containing the RF precoders and baseband precoders of each user, respectively, and are expressed as

$$\mathbf{V}_{\mathrm{RF}} = \begin{bmatrix} \mathbf{V}_{\mathrm{RF}}^{1} & \ldots & \mathbf{0} \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \ldots & \mathbf{V}_{\mathrm{RF}}^{K} \end{bmatrix}, \tag{5.3}$$

and

$$\mathbf{V}_{\mathrm{BB}} = \begin{bmatrix} \mathbf{V}_{\mathrm{BB}}^{1} & \ldots & \mathbf{0} \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \ldots & \mathbf{V}_{\mathrm{BB}}^{K} \end{bmatrix}. \tag{5.4}$$

The signal received by user k can be written as

$$\mathbf{y}_k = \sqrt{P_k}\mathbf{H}_k\mathbf{V}_{\mathrm{RF}}\mathbf{V}_{\mathrm{BB}}^{k}\mathbf{s}_k + \mathbf{H}_k \sum_{j=1, j\neq k}^{K} \sqrt{P_j}\mathbf{V}_{\mathrm{RF}}\mathbf{V}_{\mathrm{BB}}^{j}\mathbf{s}_j + \mathbf{n}_k, \tag{5.5}$$

where $\mathbf{V}_{\mathrm{BB}}^{k}$ denotes the baseband precoder for user k obtained using the its estimated channel $\hat{\mathbf{H}}_k$, and $\mathbf{n}_k \sim \mathcal{CN}(0, \sigma^2)$ is the AGWN.

5.2.3 Rate-splitting hybrid beamforming

The RS transmission strategy operates such that the message intended for each user is split into two parts: the private part and the common part. The common part of each user's message is drawn from a public codebook accessible to all users in the network, then all the common parts are put together in a common message, which can be decoded by all users with zero error probability. On the other hand, the private part of each user's message is transmitted using a fraction of the total power, and will be decoded only by the selected user. Consequently, the transmitted signal from the BS is a superimposition of the private messages intended for all K users over the common message, expressed as

$$\mathbf{x} = \sqrt{P_c}\mathbf{V}_{\mathrm{RF}}\mathbf{V}_{\mathrm{c}}\mathbf{s}_{\mathrm{c}} + \sum_{k=1}^{K}\sqrt{P_k}\mathbf{V}_{\mathrm{RF}}\mathbf{V}_{\mathrm{p}}^{k}\mathbf{s}_k, \tag{5.6}$$

where $\mathbf{s}_{\mathrm{c}}$ is the common message, and $\mathbf{s}_k$ is the private message for user k. P_k denotes the transmit power of the private message, while P_c represents the power allocated to the common message, which is the residual power after the subtraction of the power allocated to the private messages from the total transmit power.

The basic principle in this RS approach is to ensure that the fraction of power used to transmit the private message is such that they can be decoded by each user in a non-interference-limited SNR region. This guarantees that the residual power used for transmitting the common message yields a rate enhancement [131]. In the decoding procedure, first the common message is decoded by each user via SIC by treating all the private messages as noise. Then, after the removal of the common message from the received signal, each user decodes its unique private message.

5.2.4 Hardware impairment consideration

Our analysis is concerned with improving the performance degradation in mmWave systems when the available CSI is imperfect. However, the system's performance is also degraded by transceiver hardware impairments which are unavoidable in practice. Hence, in this subsection, we take into account the hardware impairments affecting the transmitters and receivers of practical mmWave systems. In particular, we present the models describing the transmitter and receiver dis-

tortion noises, respectively. For ease of exposition, we assume that hardware impairments are known by the operator through the hardware specification.

5.2.4.1 Transmitter hardware impairments

The transmitter distortion noise emerges from the combined effects of non-linearities in the power amplifier, imperfect compensation of the quantization noise in the DACs, phase noise from the local oscillators and additive power-amplifier noise [134]. Hence, this distortion models the impact of the limited dynamic range of the transmitter and is statistically independent of the transmitted signal. Therefore, the distortions at the transmitter can be modelled as [135]

$$\boldsymbol{\eta}_k \sim \mathcal{CN}\Big(\mathbf{0}, \kappa \text{diag}\left(\mathbf{V}_{\text{RF}}^k \mathbf{V}_{\text{BB}}^k (\mathbf{V}_{\text{BB}}^k)^H (\mathbf{V}_{\text{RF}}^k)^H\right)\Big), \;\; \boldsymbol{\eta}_k \perp \mathbf{x}_k, \tag{5.7}$$

where $\perp$ denotes statistical independence, and $\kappa \ll 1$ is a constant of proportionality depicting the severity of the impairments at the transmitter.

5.2.4.2 Receiver hardware impairments

Similarly, the receiver distortion captures the effect of the limited dynamic range at the receiver. These impairments encompass the non-linearities in the ADC, distortions in the local oscillator at the receiver and gain control noise. Regarding the mathematical description of the receiver noise, if $\boldsymbol{\Upsilon}_k$ is the covariance matrix of the undistorted received signal at the user, then the covariance matrix of the receiver distortion is given by β times the energy of the undistorted signal. Thus, the receiver distortion, statistically independent of the received signal, is modelled as

$$\boldsymbol{\xi}_k \sim \mathcal{CN}\Big(\mathbf{0}, \beta \, \text{diag}\left(\boldsymbol{\Upsilon}_k\right)\Big), \;\; \boldsymbol{\xi}_k \perp \mathbf{u}_k, \tag{5.8}$$

where $\mathbf{u}_k = \mathbf{y}_k - \boldsymbol{\xi}_k$ is the undistorted received signal, $\boldsymbol{\Upsilon}_k = \text{cov}\{\mathbf{u}_k\}$ and $\beta \ll 1$ as the constant of proportionality denoting the severity of the receiver impairments.[1] In other words, β is defined as the ratio between the receiver distortion noise variance and the signal power.

[1]Note that, in practical systems, the constants β and κ appear as error vector magnitudes (EVM) at the receiver and transmitter, respectively [136].

5.3 SINR and Rate Analysis with Imperfect CSI and Hardware Impairments

In this section, combining the effects of imperfect CSI with transceiver impairments, we elaborate on the performance of both conventional and RS transmission strategies in terms of sum rate and signal-to-interference-noise-ratio (SINR).

5.3.1 Conventional transmission

Under the assumption of uniform power allocation amongst all users in the system, the received signal at user k, after incorporating the aforementioned transmitter and receiver distortions, is given by

$$\mathbf{y}_k = \sqrt{P_k}\tilde{\mathbf{H}}_k\mathbf{V}_{\mathrm{RF}}\mathbf{V}_{\mathrm{BB}}^k\mathbf{s}_k + \underbrace{\tilde{\mathbf{H}}_k\sum_{j=1,j\neq k}^{K}\sqrt{P_j}\mathbf{V}_{\mathrm{RF}}\mathbf{V}_{\mathrm{BB}}^j\mathbf{s}_j}_{\text{multi-user interference}} + \mathbf{m}_k, \tag{5.9}$$

where $\tilde{\mathbf{H}}_k = [\mathbf{H}_k\ \mathbf{0}]$ is an $N_{\mathrm{R}} \times KN_{\mathrm{T}}$ matrix containing the $N_{\mathrm{R}} \times N_{\mathrm{T}}$ channel matrix $\mathbf{H}_k$ and $N_{\mathrm{R}} \times N_{\mathrm{T}}(K-1)$ appended zeros,[2] and $\mathbf{m}_k$ represents the combined contribution of transmitter and receiver distortions and the AWGN, given as

$$\mathbf{m}_k = \sqrt{P_k}\tilde{\mathbf{H}}_k\boldsymbol{\eta}_k + \sum_{j=1,j\neq k}^{K}\sqrt{P_j}\tilde{\mathbf{H}}_k\boldsymbol{\eta}_j + \boldsymbol{\xi}_k + \mathbf{n}_k. \tag{5.10}$$

From equation (5.9), the SINR of user k is expressed as

$$\mathrm{SINR}_k^{\mathrm{noRS}} = \frac{P_k\|\tilde{\mathbf{H}}_k\mathbf{V}_{\mathrm{RF}}\mathbf{V}_{\mathrm{BB}}^k\|_F^2}{P_j\sum\limits_{j=1,j\neq k}^{K}\|\tilde{\mathbf{H}}_k\mathbf{V}_{\mathrm{RF}}\mathbf{V}_{\mathrm{BB}}^j\|_F^2 + \|\boldsymbol{\Sigma}_k\|_F}, \tag{5.11}$$

[2]We note that $\mathbf{H}_k$ and $\tilde{\mathbf{H}}_k$ can be considered as equivalent since the Frobenius norms of both matrices yield the same result. Moreover, $\tilde{\mathbf{H}}_k$ is an extended version of $\mathbf{H}_k$.

where $\mathbf{\Sigma}_k$ is the covariance matrix of $\mathbf{m}_k$ approximated as[3]

$$\begin{aligned}\mathbf{\Sigma}_k \;\approx\; & \kappa\, P_k \tilde{\mathbf{H}}_k \mathrm{diag}\left(\mathbf{V}_{\mathrm{RF}} \mathbf{V}_{\mathrm{BB}}^k (\mathbf{V}_{\mathrm{BB}}^k)^H \mathbf{V}_{\mathrm{RF}}^H\right) \tilde{\mathbf{H}}_k^H \\ & + \beta\, P_k \mathrm{diag}\left(\tilde{\mathbf{H}}_k \mathbf{V}_{\mathrm{RF}} \mathbf{V}_{\mathrm{BB}}^k (\mathbf{V}_{\mathrm{BB}}^k)^H \mathbf{V}_{\mathrm{RF}}^H \tilde{\mathbf{H}}_k^H\right) \\ & + \sigma^2 \mathbf{I}\ (\beta + 1) \\ & + \sum_{j=1, j\neq k}^{K} \kappa\, P_j \tilde{\mathbf{H}}_k \mathrm{diag}\left(\mathbf{V}_{\mathrm{RF}} \mathbf{V}_{\mathrm{BB}}^j (\mathbf{V}_{\mathrm{BB}}^j)^H \mathbf{V}_{\mathrm{RF}}^H\right) \tilde{\mathbf{H}}_k^H \\ & + \sum_{j=1, j\neq k}^{K} \beta\, P_j \mathrm{diag}\left(\tilde{\mathbf{H}}_k \mathbf{V}_{\mathrm{RF}} \mathbf{V}_{\mathrm{BB}}^j (\mathbf{V}_{\mathrm{BB}}^j)^H \mathbf{V}_{\mathrm{RF}}^H \tilde{\mathbf{H}}_k^H\right). \end{aligned} \tag{5.12}$$

Given (5.11), the corresponding rate of user k is given by

$$R_k^{\mathrm{noRS}} = \log_2\left(1 + \mathrm{SINR}_k\right), \tag{5.13}$$

and the sum rate of the system is given by

$$R_{\mathrm{sum}}^{\mathrm{noRS}} = \sum_{k=1}^{K} R_k^{\mathrm{noRS}}. \tag{5.14}$$

5.3.2 RS transmission

In the RS transmission strategy, the transmit power is split between common and private messages. Assuming equal power allocation among the private messages, the received signal at user k when considering transceiver hardware impairments is defined as

$$\mathbf{y}_k = \underbrace{\sqrt{P_c}\tilde{\mathbf{H}}_k \mathbf{V}_{\mathrm{RF}} \mathbf{V}_{\mathrm{c}} \mathbf{s}_{\mathrm{c}}}_{\text{common message}} + \underbrace{\sum_{k=1}^{K} \sqrt{P_k}\tilde{\mathbf{H}}_k \mathbf{V}_{\mathrm{RF}} \mathbf{V}_p^k \mathbf{s}_{\mathrm{p}}^k}_{\text{private messages}} + \mathbf{m}_k, \tag{5.15}$$

where $P_k = \frac{P\tau}{K}$ denotes the power allocated to the private messages with τ as a fraction of the total transmit power (P), $P_c = P(1-\tau)$ represents the residual power that is allocated to the common message and $\mathbf{s}_{\mathrm{c}}$ and $\mathbf{s}_{\mathrm{p}}^k$ are the data symbol vectors for common and private messages, respectively. Furthermore, $\mathbf{V}_{\mathrm{c}}$ and $\mathbf{V}_p^k$ denotes the precoders for the common and private messages, respectively, and

[3]The approximation is obtained by omitting the terms containing the multiplication of κ and β since their product is very small and negligible.

$\mathbf{m}_k$ represents the combined noises as defined in equation (5.10).

Using equation (5.15), the SINR for the common message at user k can be computed by treating all the private messages as noise, and is given by

$$\mathrm{SINR}_{\mathrm{c}}^{k} = \frac{P_{\mathrm{c}} \|\tilde{\mathbf{H}}_k \mathbf{V}_{\mathrm{RF}} \mathbf{V}_{\mathrm{c}}\|_F^2}{\sum_{k=1}^{K} P_k \|\tilde{\mathbf{H}}_k \mathbf{V}_{\mathrm{RF}} \mathbf{V}_{\mathrm{p}}^{k}\|_F^2 + \|\mathbf{\Sigma}_k\|_F}, \tag{5.16}$$

where $\mathbf{\Sigma}_k$ is the covariance matrix of $\mathbf{m}_k$ as defined by equation (5.12). After subtracting the common message, the SINR for the private message at user k is given by

$$\mathrm{SINR}_{\mathrm{p}}^{k} = \frac{P_k \|\tilde{\mathbf{H}}_k \mathbf{V}_{\mathrm{RF}} \mathbf{V}_{\mathrm{p}}^{k}\|_F^2}{\underbrace{\sum_{j=1, j\neq k}^{K} P_j \|\tilde{\mathbf{H}}_k \mathbf{V}_{\mathrm{RF}} \mathbf{V}_{\mathrm{p}}^{j}\|_F^2}_{\text{multi-user interference}} + \|\mathbf{\Sigma}_k\|_F}. \tag{5.17}$$

In order to guarantee that the common message be decoded by all users with zero error probability, the achievable rate of the common message is determined by the user with the weakest channel gain and is given by

$$R_c = \min_k \{R_c^k\} = \min_k \{\log_2(1 + \mathrm{SINR}_{\mathrm{c}}^{k})\}. \tag{5.18}$$

On the other hand, the sum rate of all private messages is given by

$$R_{\mathrm{p}} = \sum_{k=1}^{K} R_{\mathrm{p}}^{k} = \sum_{k=1}^{K} \{\log_2(1 + \mathrm{SINR}_{\mathrm{p}}^{k})\}. \tag{5.19}$$

Accordingly, the RS sum rate is given as $R_{\mathrm{RS}} = R_{\mathrm{c}} + R_{\mathrm{p}}$.

5.3.3 Hybrid precoder design with imperfect CSI

In this subsection, we elaborate on the hybrid precoders design under imperfect channel knowledge. First, with the goal of maximising the received signal power on each BS–UE link, the analog beamformers are designed. Secondly, the digital precoders are designed for MUI nulling.

5.3.3.1 RF beamformer design

The objective in designing the RF precoder from the BS to any user is to maximise the received signal power on that link. In particular, the BS searches for beams from a feasible finite-sized RF beam-steering codebook, then the selected user feedbacks the index of the codewords that yield the maximum received signal power [80].

$\mathcal{V}$ denotes the RF beamforming codebook with low-resolution phase shifters and a large number of antenna elements. The cardinality of $\mathcal{V}$ is given by $|\mathcal{V}| = 2^{B_{\mathrm{RF}}}$, where B_{RF} denotes the number of quantization bits for the phase-shifting resolution. Consequently, for the transmission link between the BS and user k, the problem providing the optimal precoder $\mathbf{V}_{\mathrm{RF}}^{k}$ takes the form

$$\begin{aligned} &\arg\max_{\mathbf{V}_{\mathrm{RF}}^{k}} \ \|\hat{\mathbf{H}}_k \mathbf{V}_{\mathrm{RF}}^{k}\|_F^2 \\ &\text{s.t. } \ \mathbf{V}_{\mathrm{RF}}^{k} \in \mathcal{V}_{\mathrm{RF}}. \end{aligned} \tag{5.20}$$

The problem described by equation (5.20) is a typical RF beamforming design problem and can be solved using the efficient beam-training algorithms in [19, 137]. When the BS has computed the RF precoder for every user, then $\mathbf{V}_{\mathrm{RF}}$ from equation (5.3) can be computed.

5.3.3.2 Baseband precoder design

Given that $\hat{\mathbf{H}}_k$ and $\mathbf{V}_{\mathrm{RF}}$ are now known at the BS, the baseband precoder for user k can be designed using a simple multi-user beamforming strategy like regularised zero-forcing (RZF). Considering the transmitter and receiver noises, $\mathbf{V}_{\mathrm{BB}}^{k}$ is given as [138]

$$\mathbf{V}_{\mathrm{BB}}^{k} = \tilde{\mathbf{G}}_k^H \Big(\kappa\, \tilde{\mathbf{G}}_k \tilde{\mathbf{G}}_k^H + \beta\, \mathrm{diag}\left(\tilde{\mathbf{G}}_k \tilde{\mathbf{G}}_k^H \right) + \mathbf{Z}_k + N_{\mathrm{R}} \psi \mathbf{I} \Big), \tag{5.21}$$

where $\tilde{\mathbf{G}}_k = \tilde{\mathbf{H}}_k \mathbf{V}_{\mathrm{RF}}$, $\psi > 0$ is the regularization parameter scaled by N_{R}, and $\mathbf{Z}_k \in \mathbb{C}^{N_{\mathrm{R}} \times N_{\mathrm{R}}}$ is an arbitrary Hermitian positive definite matrix.[4]

5.3.4 RS precoder design

The RS transmission strategy with hybrid precoding requires the design of three precoders, namely, the RF precoder, the precoder of the private message and the

[4]Both $\mathbf{Z}_k$ and ψ can be optimised using interior point methods [139, Theorem 6] .

precoder of the common message. With $\hat{\mathbf{H}}_k$ known as the BS, the RF precoder for user k is designed through the iterative process of information exchange between the BS and user k as discussed in section 5.3.3.1. Furthermore, once $\mathbf{V}_{\mathrm{RF}}$ is designed, the precoder of the private message can be designed using a multi-user beamforming strategy and takes the form of RZF as in equation (5.21).

Regarding the design of the common precoder $\mathbf{V}_{\mathrm{c}}$, we aim to maximise the achievable rate of the common message. Since the rate of the common message is limited to the weakest user, $\mathbf{V}_{\mathrm{c}}$ can be optimised by solving the max–min problem formulated as

$$\max_{\mathbf{F_c}} \min_{k} \; q_k \, \|\tilde{\mathbf{H}}_k \mathbf{V}_{\mathrm{RF}} \mathbf{V}_{\mathrm{c}}\|_F^2 \tag{5.22a}$$

$$\text{s.t.} \qquad \mathrm{tr}\{\mathbf{V}_{\mathrm{c}}^H \mathbf{V}_{\mathrm{RF}}^H \mathbf{V}_{\mathrm{RF}} \mathbf{V}_{\mathrm{c}}\} \leq 1, \tag{5.22b}$$

where (5.22b) is the power constraint and

$$q_k = \frac{P_{\mathrm{c}}}{\sum\limits_{k=1}^{K} P_k \|\tilde{\mathbf{H}}_k \mathbf{V}_{\mathrm{RF}} \mathbf{V}_{\mathrm{p}}^k\|_F^2 + \|\mathbf{\Sigma}_k\|_F}. \tag{5.23}$$

By the introduction of an auxiliary variable t, and the use of epigraph form, the max–min optimisation problem (5.22a)–(5.22b) can be rewritten as

$$\max_{\mathbf{F_c}, t} \; t \tag{5.24a}$$

$$\text{s.t.} \; \mathrm{tr}\{\mathbf{V}_{\mathrm{c}}^H \mathbf{V}_{\mathrm{RF}}^H \mathbf{V}_{\mathrm{RF}} \mathbf{V}_{\mathrm{c}}\} \leq 1, \tag{5.24b}$$

$$t \leq q_k \, \mathrm{tr}\{\mathbf{V}_{\mathrm{c}}^H \mathbf{V}_{\mathrm{RF}}^H \tilde{\mathbf{H}}_k^H \tilde{\mathbf{H}}_k \mathbf{V}_{\mathrm{RF}} \mathbf{V}_{\mathrm{c}}\} \quad \forall \, k, \tag{5.24c}$$

where equation (5.24c) follows from the identity $\|\mathbf{A}\|_F = \sqrt{\mathrm{tr}\{\mathbf{A}\mathbf{A}^H\}}$. Replacing $\mathbf{V}_{\mathrm{c}}$ with $\mathbf{X}$, the problem described by equations (5.24a)–(5.24c) can be written as

$$\max_{\mathbf{X}, t} \; t \tag{5.25a}$$

$$\text{s.t.} \; \mathrm{tr}\{\mathbf{X}^H \mathbf{V}_{\mathrm{RF}}^H \mathbf{V}_{\mathrm{RF}} \mathbf{X}\} \leq 1, \tag{5.25b}$$

$$t \leq q_k \, \mathrm{tr}\{\mathbf{X}^H \mathbf{V}_{\mathrm{RF}}^H \tilde{\mathbf{H}}_k^H \tilde{\mathbf{H}}_k \mathbf{V}_{\mathrm{RF}} \mathbf{X}\} \quad \forall \, k, \tag{5.25c}$$

Note that the problem described by equations (5.24a)–(5.24c) is non-convex due to coupling of optimisation variables and the non-convex constraint (5.24c).

Hence, it is difficult to obtain a tractable optimal solution in its current state. Accordingly, we transform the problem by applying the $vec(.)$ operation, the use of the identity $\|vec(\mathbf{A})\|_2^2 = \mathrm{tr}\{\mathbf{A}\mathbf{A}^H\}$, and introducing a new variable $\mathbf{Z}$ such that for any $\mathbf{Z}$, we have that

$$(\mathbf{X}-\mathbf{Z})^H \mathbf{H}_{k,\mathrm{EFF}}^H \mathbf{H}_{k,\mathrm{EFF}} (\mathbf{X}-\mathbf{Z}) \geq 0, \tag{5.26}$$

where $\mathbf{H}_{k,\mathrm{EFF}} = \tilde{\mathbf{H}}_k \mathbf{V}_{\mathrm{RF}}$. In this regard, the maximisation optimisation problem (5.25a)–(5.25c) can be written as

$$\max_{\mathbf{X},t} \; t \tag{5.27a}$$

$$\text{s.t.} \; \|vec(\mathbf{V}_{\mathrm{RF}}\mathbf{X})\|_2^2 \leq 1, \tag{5.27b}$$

$$q_k \,\mathrm{tr}\{2\,\mathrm{Re}\left[\mathbf{Z}^H \mathbf{H}_{k,\mathrm{EFF}}^H \mathbf{H}_{k,\mathrm{EFF}} \mathbf{X}\right] - \mathbf{Z}^H \mathbf{H}_{k,\mathrm{EFF}}^H \mathbf{H}_{k,\mathrm{EFF}} \mathbf{Z}\} \geq t \;\; \forall\, k, \tag{5.27c}$$

where

$$\mathbf{X}^H \mathbf{H}_{k,\mathrm{EFF}}^H \mathbf{H}_{k,\mathrm{EFF}} \mathbf{X} \;\geq\; 2\,\mathrm{Re}\left[\mathbf{Z}^H \mathbf{H}_{k,\mathrm{EFF}}^H \mathbf{H}_{k,\mathrm{EFF}} \mathbf{X}\right] - \mathbf{Z}^H \mathbf{H}_{k,\mathrm{EFF}}^H \mathbf{H}_{k,\mathrm{EFF}} \mathbf{Z}. \tag{5.28}$$

The problem described by equations (5.27a)–(5.27c) is a second-order cone programming (SOCP) problem since the objective function (5.27a) is linear and the constraints (5.27b)–(5.27c) are second-order cones. Accordingly, equations (5.27a)–(5.27c) can be efficiently solved through an iterative maximisation process with standard second-order cone programming (SOCP) solvers as outlined in Algorithm 5.1.

Algorithm 5.1: Alternating maximisation process to solve common precoder.

1. **Require** $\hat{\mathbf{H}}_k$ and $\mathbf{V}_{\mathrm{RF}}$
2. Set the iteration number $n = 0$ and initialize $\mathbf{Z}$
3. $n \leftarrow n + 1$.
4. Obtain $\mathbf{X}^*$ by solving (5.27a)–(5.27c)
5. Update $\mathbf{Z} = \mathbf{X}^*$
6. Repeat from step 3 until convergence or a fixed number of iterations is reached.
7. Return $\mathbf{V}_{\mathrm{c}} = \mathbf{X}^*$.

5.3.5 Power allocation

To determine the power allocation in the RS transmission strategy for each user accurately, the saturation point of the achievable rate of the private message is first determined. In other words, we evaluate the point at which the multi-user interference becomes the dominant factor in the achievable rate of the private message. At this point, increasing the power allocated to the private message has no tangible effect and power saturation occurs. Once the power saturation point has been determined, the residual power is allocated to the common message.

Mathematically, the fraction of transmit power allocated to the private message $\tau \in (0, 1]$ for user k is expressed as [131]

$$\tau = \min\{\tfrac{K}{P\Upsilon}, 1\}, \tag{5.29}$$

where $\Upsilon = \min_k\{\frac{1}{K}\sum_{j\neq k}\|\tilde{\mathbf{H}}_k\mathbf{V}_{\mathrm{RF}}\mathbf{V}_{\mathrm{p}}^j\|_F^2\}$ represents the interference power from other users in the network.

Remark 5.1. *The case of $\tau = 0$ implies that the common part is transmitted with the full transmit power resulting in an* RS *rate limited by the* UE *with the weakest channel. Accordingly, this case ($\tau = 0$) is meaningless and excluded from the analysis. On the other hand, the case of $\tau = 1$ indicates that full transmit power is allocated to the private messages and as a result the* RS *rate of this case ($\tau = 1$) is equivalent to the rate of conventional transmission strategy in equation* (5.14).

5.4 Numerical Results

In this section, we evaluate the performance of the proposed RS transmission scheme in a multi-user mmWave system. We consider a mmWave system operating at 28-GHz carrier frequency with a bandwidth of 100 MHz where a BS communicates with 4 UEs. Unless otherwise stated, the BS has $N_{\mathrm{T}} = 64$ antennas and $N_{\mathrm{T}}^{\mathrm{RF}} = 16$ RF chains, while each UE has $N = 16$ antennas and $N_{\mathrm{T}}^{\mathrm{RF}} = 8$ RF chains, and the transmitter and receiver distortions are $\kappa = \beta = -60$ dB [120].

Throughout the simulations, RF beamforming vectors are made up of elements of quantized phase shifts. Thus, we implement the beamforming vectors with 7

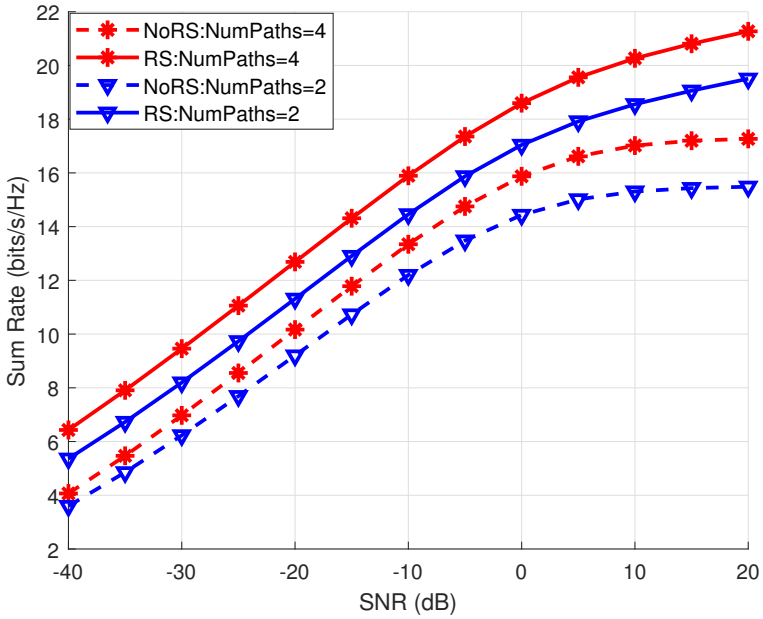


Figure 5.2: Comparison of the sum-rate performance achieved by RS and NoRS transmission strategies for multi-path channels with resolution parameter $N_{\text{res}} = 256$ and $Q = 2$ beamforming training vectors.

quantization bits. For channel estimation, we employ the method based on hybrid architecture detailed in section 2.4.2 with the sparse channel formulation given as $\mathbf{y} = \sqrt{P}\left(\mathbf{V}_t^T \otimes \mathbf{W}_t^H\right)\mathbf{A}_D\mathbf{h}_\alpha + \mathbf{z}$, where $\mathbf{z}$ represents the additive noise vector. Accordingly, the channel parameters are obtained by estimating the locations of the non-zero elements of $\mathbf{h}_\alpha$ which correspond to the AoAs/AoDs and the channel path gains. In particular, adopting the adaptive CS algorithm in [60], the AoA/AoDs are estimated by means of a multi-level resolution codebook with resolution parameter N_{res} and with Q beamforming training vectors for each level. We employ equal transmit power P in all precoding and combining solutions. Note that P is dependent on SNR, i.e., SNR$= \frac{P}{\sigma^2}$, where σ^2 denotes the noise variance.

Our aim is to shed light on the performance of RS in a pragmatic mmWave system where imperfect CSI is taken into account. Hence, throughout the simulations we provide a comparison of the conventional NoRS transmission strategy with the proposed RS approach. Specifically, in Figs. 5.2–5.6, the solid lines depict the simulations of the RS method while the corresponding dotted lines correspond to the NoRS cases. In Fig. 5.2, we analyse the impact of scattering and blockages that contribute to propagation paths in mmWave systems. Conse-

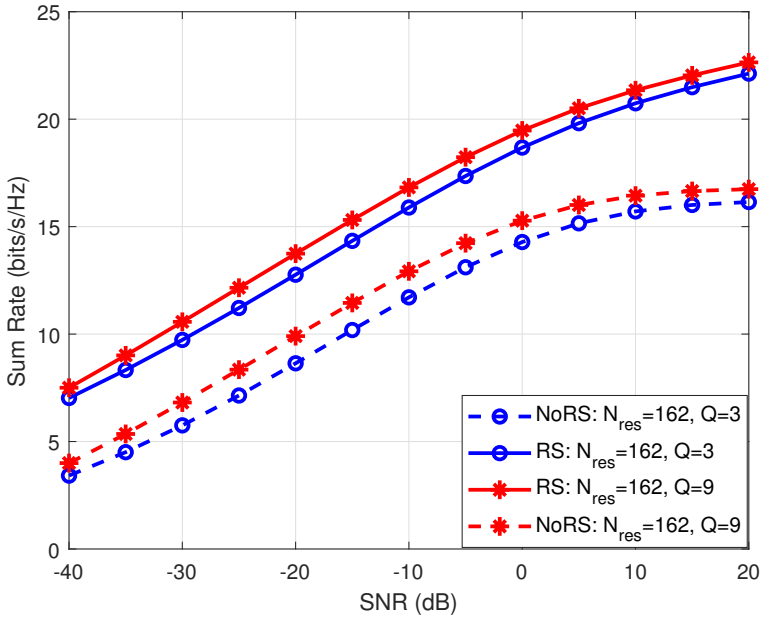


Figure 5.3: Comparison of the sum-rate performance achieved by RS and NoRS transmission strategies for multi-path channels with varying beamforming vectors and resolution parameter $N_{\text{res}} = 162$.

quently, Fig. 5.2 illustrates the sum-rate performance achieved by the estimation of multi-path channels. It can be seen that rate gain is increased for a higher number of paths since the hybrid architecture is implemented with the number of multiplexed streams equal to the number of paths. In addition, it is quite evident that, in all cases, the RS transmission method achieves a much higher sum rate than the NoRS schemes. For example, when the two-path channels are estimated, the gain of the RS strategy over the corresponding NoRS is about 4.5 bits/s/Hz at 0 dB. This validates the effectiveness of the proposed RS scheme for multi-path mmWave channels.

The accuracy of the channel estimation when using adaptive CS methods is dependent on the multi-resolution codebook structure used to estimate the channel and beamforming vectors i.e., the resolution parameter N_{res} and the Q number of beamforming vectors at each codebook level. Hence, in Figs 5.3 and 5.4 we analyse the impact of varying the number of beamforming vectors and the resolution parameter, respectively. We set the number of paths to 2 and the resolution parameter $N_{\text{res}} = 162$ in Fig. 5.3. It can be observed that the rate gain for both transmission schemes is decreased when a lower number of beamforming vectors ($Q = 3$) are used for training the RF precoders/combiners. However,

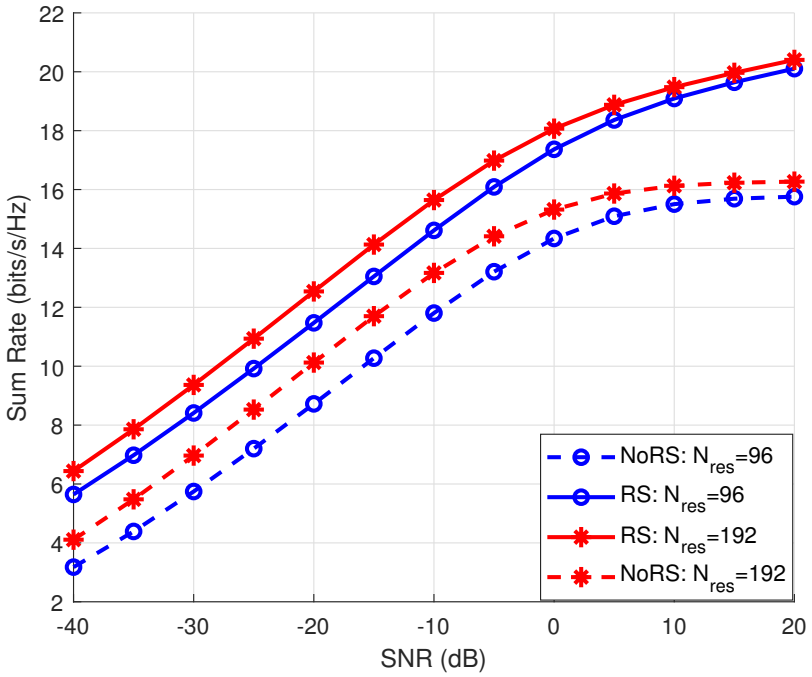


Figure 5.4: Comparison of the sum-rate performance achieved by RS and NoRS transmission strategies for multi-path channels with varying resolution parameters and $Q = 2$ beamforming training vectors.

it should be noted that the beamforming vectors impact the training overhead and should not be increased indefinitely. Further, the RS transmission strategy performs better than the NoRS scheme, suggesting that even in low-complexity scenarios, RS can achieve comparable gains. In Fig. 5.4, the number of paths are set to 3, it is evident that the quantization loss decreases for increasing resolution parameters. The results from both Figs 5.3 and 5.4 suggest that the RS transmission scheme always provides an improvement in the sum-rate performance over the NoRS strategies regardless of the available channel knowledge.

Next, in Fig. 5.5, we illustrate the comparison of the sum-rate performance between employing the RS transmission strategy and the NoRS schemes in a mmWave system with respect to the transmitter and receiver distortions. For brevity, we assume equal transmitter and receiver distortions, i.e., $\kappa = \beta$. We set $N_{\text{res}} = 96$, $Q = 2$, and the SNR to 20 dB for estimating three-path channels. It can be seen from Fig. 5.5 that when the distortion in the transceiver is low for both transmission schemes, the achievable sum rate is high. For instance, at $\kappa = \beta = -60$, a sum rate of 22 bits/s/Hz is achieved by using the RS transmission scheme while the corresponding rate achieved by the NoRS approach is

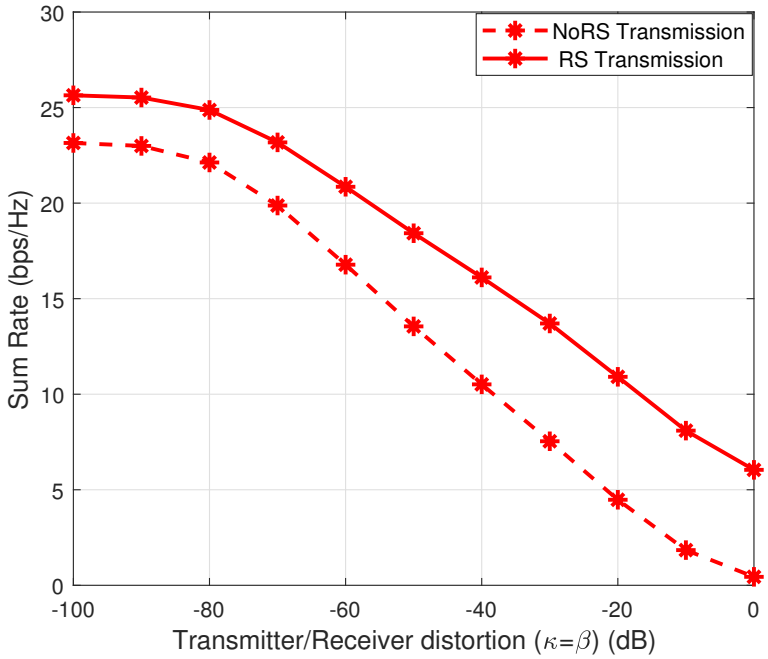


Figure 5.5: Achievable sum rate of RS and NoRS transmission strategies with respect to transmitter and receiver distortions.

17 bits/s/Hz. However, the rate decreases steadily as the transceiver distortions increase. This suggests that the transmitter/receiver distortions affect the performance of system and should be considered in the design of the precoders and combiners. Notwithstanding, the case of RS is desirable, showing that splitting the power between a common and private message achieves additional gain to counter the degradation caused by the hardware impairments.

Finally, we investigate the impact of varying the number of users on the sum-rate performance of the mmWave system in Fig. 5.6. It is shown clearly that, with both transmission schemes, increasing the number of UEs leads to a decrease in the sum-rate performance. This can be explained by the fact that there is increased MUI in the system resulting from more interactions by more users. Specifically, at 20 dB in the NoRS transmission case, when the BS communicates to 3 UEs, the sum rate achieved is approximately 24 bits/s/Hz, attaining a 30% improvement in sum-rate gain over approximately 21 bits/s/Hz obtained when there are 4 users. Nevertheless, the RS transmission scheme is even more desirable in both scenarios; this is as expected since, by power-splitting in the RS transmission, residual MUI due to CSI mismatch is tackled. This further validates the effectiveness of applying RS strategies to multi-user mmWave systems.

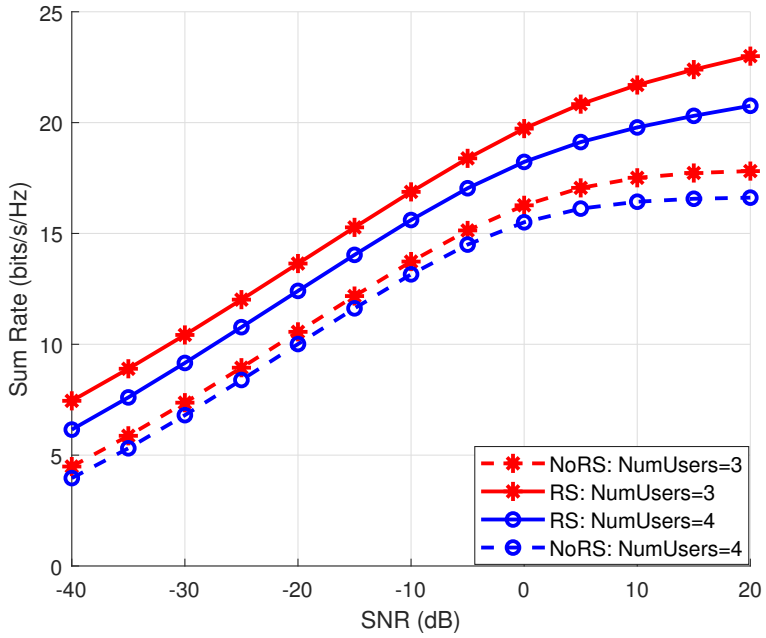


Figure 5.6: Comparison of the sum-rate performance achieved by RS and NoRS transmission strategies for varying number of users with resolution parameter $N_{\text{res}} = 96$ and $Q = 2$ beamforming training vectors.

5.5 Conclusion

In multi-user mmWave networks, imperfect CSI results in MUI, which adversely affects the sum-rate performance of the system. Accordingly, this chapter investigated the RS transmission strategy for multi-user mmWave communications with hybrid precoding. The estimated channel was derived from estimates of the angles of departure, arrival and gain of each path, and the hybrid beamforming matrices were constructed using the multi-resolution codebook structure. The numerical results showed that the RS transmission scheme is beneficial to multi-user mmWave systems with imperfect CSI, as the residual MUI in the system, caused due to channel estimation errors, is tackled by the power splitting in RS. Additionally, it was shown that increasing the resolution parameter increases the efficiency of the RS scheme. Moreover, the RS strategy is robust for practical scenarios with constrained feedback overheads and imperfect hardware.

Chapter 6

Impact of Hardware Impairments on mmWave MIMO Systems

6.1 Introduction

The research background for hybrid beamforming in mmWave systems has been presented in section 2.3.2, along with a description of some of the most recent techniques that may contribute to the improvement of the system's performance. Among the classes of hybrid beamforming techniques, the structures implemented with digitally controlled phase shifters have been identified as one of the promising measures that may be used with quantized phases to correct the lack of precision in analog beamforming techniques [28, 43, 42, 81, 140, 141].

A convention among these studies regarding hybrid beamforming in mmWave MIMO communications is the assumption of ideal hardware. In practice, however, the RF front-end (where amplification and down conversion of signals occur) operates at dynamic ranges of frequencies which are much higher than the operating frequencies of the baseband [134]. On this ground, the signal processing at baseband level cannot remove the hardware impairments which arise from non-linearities of the amplifier, phase noise, quantization errors, mutual coupling between antenna ports, and I/Q imbalance [100, 101]. Moreover, low-power circuits, which are more susceptible to hardware impairments, are deployed by the cost-effective implementation of massive multi-antenna systems [135]. These transceiver hardware impairments present a design challenge for commercially vi-

Work in this chapter has been presented at IEEE WCNC, 2018 [102].

able mmWave MIMO transceivers. Therefore, this chapter extends existing work by investigating the impact of hardware imperfections on mmWave systems with hybrid beamforming from a design perspective.

In standard massive MIMO cellular systems operating at lower frequencies, hardware impairments and their impact have been studied extensively (e.g., see [142, 143, 144, 145, 146, 104]). Specifically, three major sources of hardware impairments are the multiplicative impairments, the amplified thermal noise (ATN), and residual additive transceiver impairments (RATHI) [143, 145, 104]. Multiplicative imperfections such as amplitude and phase errors are modelled using a stochastic error model in [143]. Using this model, the work in [146] showed that the degradation in performance due to user mobility is increased by phase errors. The additive impairments are modelled as power-dependent Gaussian additive noise [144], and results from [147] show that RATHIs have a greater impact on downlink transmissions than the uplink counterparts.

Notably, fewer studies have considered hardware impairments from a mmWave perspective. In [148], the performance of a mmWave system employing analog beamforming under hardware impairments is evaluated by using a logarithmic error model, which assumes that the combiner losses are dependent on the number of RF chains. However, analog beamforming is not robust for multi-stream or multi-user scenarios. Moreover, the other major impairments of additive distortions and amplified thermal noises are not considered in [148]. On the other hand, the work of [149] studies the major hardware impairments using othorgonal frequency division multiplexing (OFDM) waveforms. For this chapter, we move beyond the assumptions of analog beamforming [148] and digital precoding/combining (the limitations of which have been outlined earlier in section 2.3) [149] to consider a mmWave MIMO system with hybrid beamforming and provide an assessment of the impact of these key impairments on the system, with the potential to aid hybrid beamforming designs. We develop a framework for the modelling of RTHIs and, within this context, design the hybrid precoders and combiners. Additionally, we provide a detailed analysis on how the variation of key factors of the system affects each individual hardware impairment. Results show that an increase in the number of RF chains mitigates the effect of the multiplicative PN. This presents a unique observation regarding the spectral efficiency performance and should be considered in the design of mmWave systems with hybrid precoding. Furthermore, for a general overview, we provide comparisons with outcomes obtained under the assumption of ideal hardware.

The remainder of this chapter is organised as follows. Section 6.1a presents the basic parameters of the mmWave MIMO system model with hybrid precoding. In section 6.3, we provide a description of the various hardware impairments under consideration and model the spectral efficiency under the presence of these RTHIs. We provide the hybrid precoder and combiner designs over practical mmWave MIMO channels with RTHIs in section 6.4 and section 6.4.2 respectively. The numerical results are presented in section 6.5. Finally, section 6.6 provides the concluding remarks.

6.2 System Model

We begin by considering a basic mmWave MIMO system with hybrid precoding as illustrated in Fig. 6.1a, then subsequently we enrich this model by introducing (i) the multiplicative phase noise, (ii) the residual additive transceiver impairments and (iii) the amplified thermal noise, which are unavoidable in practical systems, as indicated in Fig. 6.1b.

6.2.1 Conventional model

The basic mmWave system consists of a transmitter with N_{T} antennas communicating with an N_{R}-antenna receiver by means of N_{s} data streams. Multi-stream communication is enabled if the transmitter deploys $N_{\mathrm{T}}^{\mathrm{RF}}$ transmit chains under the constraint $N_{\mathrm{s}} \leq N_{\mathrm{T}}^{\mathrm{RF}} \leq N_{\mathrm{T}}$. In other words, a hybrid hardware architecture is designed, where an $N_{\mathrm{T}}^{\mathrm{RF}} \times N_{\mathrm{s}}$ digital baseband precoder, $\mathbf{V}_{\mathrm{BB}}$, with $N_{\mathrm{T}}^{\mathrm{RF}}$ transmit chains is accompanied by an $N_{\mathrm{T}} \times N_{\mathrm{T}}^{\mathrm{RF}}$ analog precoder, $\mathbf{V}_{\mathrm{RF}}$. Considering the downlink transmission and assuming that a single carrier waveform is used, the discrete-time transmitted signal at sample time-interval n is given by

$$\mathbf{x}(n) = \mathbf{V}_{\mathrm{RF}}\, \mathbf{V}_{\mathrm{BB}}\, \mathbf{s}(n) \tag{6.1}$$

where $\mathbf{s}(n) = [s_1(n), \dots s_{N_{\mathrm{s}}}(n)]$ is the $N_{\mathrm{s}} \times 1$ symbol vector obeying $\mathbb{E}[\mathbf{s}\mathbf{s}^H] = \frac{1}{N_{\mathrm{s}}}\mathbf{I}_{N_{\mathrm{s}}}$.

When implementing $\mathbf{V}_{\mathrm{RF}}$ with analog phase shifters, all of its elements have equal norm since they satisfy $(\mathbf{V}_{\mathrm{RF}}^{(i)}\, \mathbf{V}_{\mathrm{RF}}^{(i)H})_{k,k} = \frac{1}{N_{\mathrm{T}}}$. Note that $(.)_{k,k}$ denotes the kth diagonal element of a matrix. Regarding the total power constraint of the transmitter $\mathbf{V}_{\mathrm{BB}}$, it is normalized such that $||\mathbf{V}_{\mathrm{RF}}\, \mathbf{V}_{\mathrm{BB}}||_F^2 = N_{\mathrm{s}}$. Adopting a

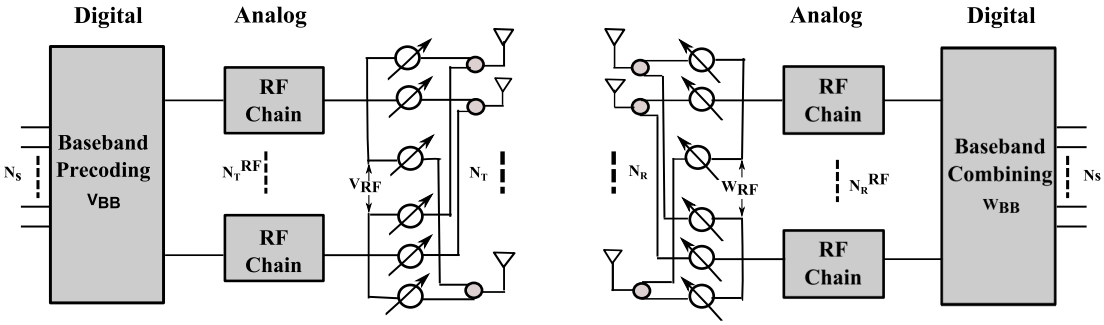


Figure 6.1a: An illustration of a basic hybrid mmWave cellular MIMO network with hybrid precoding and combining.

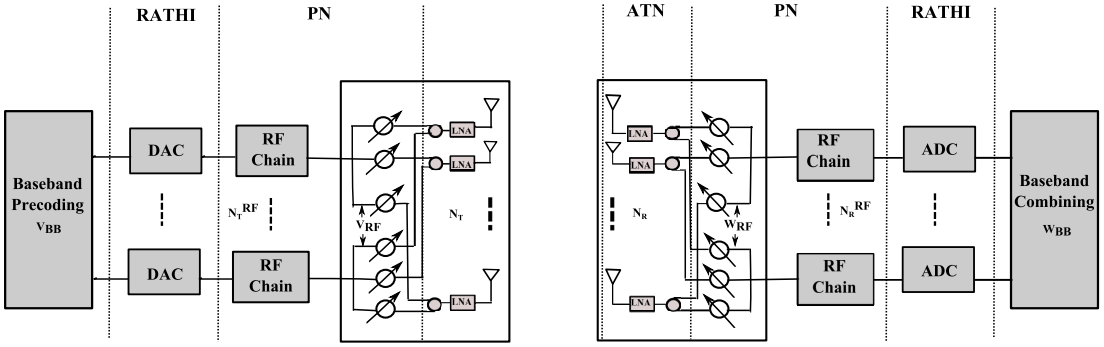


Figure 6.1b: An illustration of a realistic basic hybrid mmWave cellular MIMO network with transceiver hardware impairments.

narrowband block-fading channel model [23], the $N \times 1$ received signal is given by

$$\mathbf{y}(n) = \mathbf{H}\mathbf{V}_{\mathrm{RF}}\,\mathbf{V}_{\mathrm{BB}}\,\mathbf{s}(n) + \mathbf{z}(n), \tag{6.2}$$

where $\mathbf{z}$ is the Gaussian noise vector such that $\mathbf{z} \sim \mathcal{CN}(0, \sigma_z^2 \mathbf{I}_N)$ and $\mathbf{H}$ represents the $N \times N_{\mathrm{T}}$ channel matrix which satisfies $\mathbb{E}[||\mathbf{H}||_F^2] = N_{\mathrm{T}} N_{\mathrm{R}}$ and is expressed as

$$\mathbf{H} = \sqrt{\frac{N_{\mathrm{T}} N_{\mathrm{R}}}{N_{\mathrm{cl}} N_{\mathrm{p}}}} \sum_{l=1}^{N_{\mathrm{cl}}} \sum_{k=1}^{N_{\mathrm{p}}} \alpha_{l,k} \mathbf{a}_{\mathrm{R}} \left(\phi_{\mathrm{R}}^{l,k}, \theta_{\mathrm{R}}^{l,k}\right) \mathbf{a}_{\mathrm{T}}^{H} \left(\phi_{\mathrm{T}}^{l,k}, \theta_{\mathrm{T}}^{l,k}\right), \tag{6.3}$$

where N_{cl}, N_{p} and $\alpha_{l,k}$ denote the number of scattering clusters, the number of paths per cluster and the complex gain of the kth path in the lth cluster, respectively. $\mathbf{a}_{\mathrm{T}}$ and $\mathbf{a}_{\mathrm{R}}$ represent the transmit and receive antenna array response vectors.[1] In a similar way, the receiver uses $N_{\mathrm{R}}^{\mathrm{RF}}$ chains and analog shifters such that $N_{\mathrm{s}} \leq N_{\mathrm{R}}^{\mathrm{RF}} \leq N_{\mathrm{R}}$ for processing the received signal. Here, the processed

[1]Implicitly, this model, focusing on the transceiver hardware impairments and not on the knowledge of the channel, is based on the strong assumption that the transmitter and receiver have instantaneous and perfect knowledge of the channel $\mathbf{H}$.

$N_\mathrm{s} \times 1$ signal is given by

$$\mathbf{y}_\mathrm{s}(n) = \mathbf{W}_\mathrm{BB}^H \mathbf{W}_\mathrm{RF}^H \mathbf{H} \mathbf{V}_\mathrm{RF} \mathbf{V}_\mathrm{BB} \mathbf{s}(n) + \mathbf{W}_\mathrm{BB}^H \mathbf{W}_\mathrm{RF}^H \mathbf{z}(n), \tag{6.4}$$

where $\mathbf{W}_\mathrm{BB}$ is the $N_\mathrm{R}^\mathrm{RF} \times N_\mathrm{s}$ baseband combining matrix, and $\mathbf{W}_\mathrm{RF}$ is the $N_\mathrm{R} \times N_\mathrm{R}^\mathrm{RF}$ analog combining matrix, which is implemented using phase shifters and obeys $(\mathbf{W}_\mathrm{RF}^{(i)} \mathbf{W}_\mathrm{RF}^{(i)H})_{k,k} = \frac{1}{N_\mathrm{R}}$.

If the transmission consists of Gaussian symbols, the spectral efficiency, achieved over the aforementioned mmWave MIMO channel, is given by [150]

$$\mathcal{R} = \log_2 \left(\left| \mathbf{I}_{N_\mathrm{s}} + \frac{\rho}{N_\mathrm{s}} \mathbf{R}^{-1} \mathbf{W}^H \mathbf{H} \mathbf{V} \mathbf{V}^H \mathbf{H}^H \mathbf{W} \right| \right), \tag{6.5}$$

where ρ is the average received power and $\mathbf{R} = \sigma^2 \mathbf{W}^H \mathbf{W}$ denotes the noise covariance matrix with $\mathbf{W} = \mathbf{W}_\mathrm{RF} \mathbf{W}_\mathrm{BB}$, and $\mathbf{V} = \mathbf{V}_\mathrm{RF} \mathbf{V}_\mathrm{BB}$.

6.3 Hardware Impairments

The transceiver of practical systems is affected by various impairments, as illustrated in Fig. 6.1b. These impairments are presented below by means of their models. Fortunately, the impairments can be described physically and mathematically. In particular, we focus on (1) the multiplicative PN at both the transmitter and the receiver, (2) the residual additive power dependent distortions at the transmitter and receiver and (3) the ATN at the receiver side.

6.3.1 Multiplicative PN

Physically, the PN is defined as a distortion in the phase caused by the random phase drifts in the signal. This distortion, which comes from the local oscillators (LOs) of the transmitter and the receiver is induced during the up-conversion of the baseband signal to passband and vice-versa. Mathematically, this conversion takes place by multiplying the signal with the LO's output. In other words, every RF transceiver chain is affected by PN, expressed as

$$s_\mathrm{out}(n) = s_\mathrm{in}(n) e^{j\psi(n)}, \tag{6.6}$$

where $s_{\text{out}}(n)$ and $s_{\text{in}}(n)$ represent the output and input signal of each transceiver, and $\psi(n)$ denotes the instantaneous phase noise at the nth time interval, which can be described by a discrete-time independent Wiener process [104, 151, 152] and expressed as

$$\psi_i(n) = \psi_i(n-1) + \delta^{\psi_i(n)},\ i = T, R \tag{6.7}$$

where $\delta^{\psi_i(n)} \sim \mathcal{N}(0, \sigma^2_{\psi_i})$ denotes the PN increment variance given by $\sigma^2_{\psi_i} = 4\pi^2 f_c\, c_n\, T_s$ with f_c, c_n and T_s referring to the carrier frequency, a constant dependent on the oscillator and the symbol interval, respectively. Denoting $\mathbf{\Psi}(n) \triangleq \text{diag}\{e^{j\psi_{T,1}(n)}, \ldots, e^{j\psi_{T,N_{\text{T}}}(n)}\}$, and $\mathbf{\Omega}(n) \triangleq \text{diag}\{e^{j\psi_{R,1}(n)}, \ldots, e^{j\psi_{R,N_{\text{R}}}(n)}\}$ as the total PN matrices due to imperfections in the LOs of the transmitter and receiver, respectively. The received signal under PN is given by

$$\mathbf{y}_{\text{s}}(n) = \mathbf{W}^H_{\text{BB}}\left(\mathbf{\Omega}(n)\right)^H \mathbf{W}^H_{\text{RF}}\mathbf{H}\mathbf{V}_{\text{RF}}\mathbf{\Psi}(n)\mathbf{V}_{\text{BB}}\mathbf{s}(n) + \mathbf{W}^H_{\text{BB}}\left(\mathbf{\Omega}(n)\right)^H \mathbf{W}^H_{\text{RF}}\mathbf{z}(n). \tag{6.8}$$

6.3.2 RATHIs

The second category of hardware impairments that we are considering are the RATHIs that emerge from the DC offset, the antenna coupling, I/Q mismatch and the imperfect compensation of the quantization noise in the ADCs at the receiver and in the DACs at the transmitter [134]. To be more specific, at the transmitter, the distortions occur from a mismatch appearing between the signal that is intended to be transmitted and the generated signal, while the received signal is distorted during the reception processing at the receiver side. These impairments appear as additive variables in the signal model.

Regarding the mathematical description of the additive noises, measurement results from [101] show that the additive transmitter and receiver distortion noises are distributed as circularly-symmetric complex Gaussian random variables with their average power proportional to the signal power. This Gaussian distribution can further be justified by aggregating the contributions of many impairments [103, 105, 134]. More precisely, denoting $\mathbf{Q} \triangleq \text{diag}\{q_1, \ldots, q_{N_{\text{T}}}\}$ as the transmit covariance matrix, the distortions at the transmitter and the receiver are expressed, respectively, as

$$\boldsymbol{\eta}_{\text{T}} \sim \mathcal{CN}(\mathbf{0}, \mathbf{\Lambda}_n) \tag{6.9}$$

$$\boldsymbol{\eta}_{\text{R}} \sim \mathcal{CN}(\mathbf{0}, \mathbf{\Upsilon}_n), \tag{6.10}$$

where $\mathbf{\Lambda}_n = (\kappa_{\mathrm{T}}(n))^2 \, \mathrm{diag}\{q_1, \ldots, q_{N_{\mathrm{T}}}\}$, and $\mathbf{\Upsilon}_n = (\kappa_{\mathrm{R}}(n))^2 \, \mathbf{H}^H \mathbf{Q} \mathbf{H}$ with κ_{T}^2 and κ_{R}^2 denoting the constants of proportionality which depict the severity of the residual additive impairments at the transmitter and receiver. In other words, they are defined as the ratio between the additive distortion noise variance and the signal power. In practical systems, these constants appear as EVMs at the transceiver [136].

6.3.3 ATN

The third category of hardware impairments are the ATNs which affect the receiver. The ATN is described by an amplification that comes from the components of the receiver hardware, including the low noise amplifiers and mixers as well as leakages from other frequency bands, and is expressed by an increase in the thermal noise variance at the receiver [135]. The overall ATN impact is modelled as a Gaussian distribution, i.e, $\boldsymbol{\xi} \sim (\mathbf{0}, \xi_z \mathbf{I}_{N_{\mathrm{R}}})$, where $\xi_z \geq \sigma_z^2$, and σ_z^2 is the parameter representing the actual thermal noise.

Remark 6.1. *The ideal model, which does not account for the unavoidable hardware impairments, is obtained if* $\sigma_{\psi_i} = \kappa_{\mathrm{T}} = \kappa_{\mathrm{R}} = 0$ *and* $\xi_z = \sigma_z^2$.

6.3.4 Transmission with RTHIs

After incorporating the aforementioned transceiver impairments into the system model, the received signal becomes

$$\begin{aligned} \mathbf{y}_{s(n)} &= \mathbf{W}_{\mathrm{BB}}^H \left(\mathbf{\Omega}(n)\right)^H \mathbf{W}_{\mathrm{RF}}^H \mathbf{H} \left(\mathbf{V}_{\mathrm{RF}} \mathbf{\Psi}(n) \mathbf{V}_{\mathrm{BB}} \mathbf{s}(n) + \boldsymbol{\eta}_{\mathrm{T}}(n)\right) \\ &\quad + \mathbf{W}_{\mathrm{BB}}^H \left(\mathbf{\Omega}(n)\right)^H \mathbf{W}_{\mathrm{RF}}^H \left(\boldsymbol{\eta}_{\mathrm{R}}(n) + \boldsymbol{\xi}(n)\right). \end{aligned} \tag{6.11}$$

Within this context, the spectral efficiency is given by

$$\mathcal{R} = \log_2 \left(\left| \mathbf{I}_{N_{\mathrm{s}}} + \frac{\rho}{N_{\mathrm{s}}} \mathbf{R}^{-1} \mathbf{W}_{\Omega}^H \mathbf{H} \mathbf{V}_{\Psi} \mathbf{V}_{\Psi}^H \mathbf{H}^H \mathbf{W}_{\Omega} \right| \right), \tag{6.12}$$

where $\mathbf{W}_{\Omega} = \mathbf{W}_{\mathrm{RF}} \mathbf{\Omega}_n \mathbf{W}_{\mathrm{BB}}$, $\mathbf{V}_{\Psi} = \mathbf{V}_{\mathrm{RF}} \mathbf{\Psi} \mathbf{V}_{\mathrm{BB}}$ and the noise covariance matrix is given by

$$\mathbf{R} = \left(\kappa_{\mathrm{T}}^2 \frac{\rho}{N_{\mathrm{s}}} + \kappa_{\mathrm{R}}^2 \frac{\rho}{N_{\mathrm{s}}} \mathbf{V}_{\Psi} \mathbf{V}_{\Psi}^H \right) \mathbf{W}_{\Omega}^H \mathbf{H} \mathbf{H}^H \mathbf{W}_{\Omega} + \xi_z \mathbf{W}_{\Omega}^H \mathbf{W}_{\Omega}. \tag{6.13}$$

6.4 Hybrid Beamforming Design for a mmWave Channel with RTHIs

The focus of this section concerns the maximization of the SE expression given by equation (6.12), and the appropriate design of the hybrid mmWave precoders, $(\mathbf{V}_{\mathrm{RF}}, \mathbf{V}_{\mathrm{BB}})$, and the hybrid mmWave combiners $(\mathbf{W}_{\mathrm{RF}}, \mathbf{W}_{\mathrm{BB}})$. Hence, the SE maximization problem is formulated as

$$\max_{\mathbf{V}_{\mathrm{RF}}, \mathbf{V}_{\mathrm{BB}}, \mathbf{W}_{\mathrm{RF}}, \mathbf{W}_{\mathrm{BB}}} \mathcal{R}(\mathbf{V}_{\mathrm{RF}}, \mathbf{V}_{\mathrm{BB}}, \mathbf{W}_{\mathrm{RF}}, \mathbf{W}_{\mathrm{BB}}) \tag{6.14a}$$

$$\text{s.t.} \quad \mathbf{W}_{\mathrm{RF}} \in \mathcal{W}_{\mathrm{RF}} \tag{6.14b}$$

$$\mathbf{V}_{\mathrm{RF}} \in \mathcal{V}_{\mathrm{RF}} \tag{6.14c}$$

$$\|\mathbf{V}_{\Psi}\|_F^2 = N_{\mathrm{s}}. \tag{6.14d}$$

Note that this optimisation problem (6.14a)–(6.14d) requires the joint optimisation of four matrix variables concerning the mmWave precoders and combiners with non-convex constraints on $\mathbf{V}_{\mathrm{RF}}$ and $\mathbf{W}_{\mathrm{RF}}$. Unfortunately, an exact solution is unlikely and the corresponding problem is intractable [153]. For this reason, we follow a similar design to [28] and decouple the joint precoder–combiner optimisation problem.

6.4.1 Hybrid precoder design

With the objective of optimising the SE, we engage in the design of the hybrid precoders $(\mathbf{V}_{\mathrm{RF}}, \mathbf{V}_{\mathrm{BB}})$ maximizing the mutual information by means of Gaussian signalling over a practical mmWave MIMO channel with transceiver hardware impairments. Hence, the mutual information, dependent only on the transmitter operation, is given by

$$\mathcal{I}(\mathbf{V}_{\Psi}) = \log_2(|\mathbf{I}_{N_{\mathrm{s}}} + \frac{\rho}{N_{\mathrm{s}}} \mathbf{R}_{\mathrm{p}}^{-1} \mathbf{H} \mathbf{V}_{\Psi} \mathbf{V}_{\Psi}^H \mathbf{H}^H|), \tag{6.15}$$

where $\mathbf{R}_{\mathrm{p}} = \beta \mathbf{H}\mathbf{H}^H + \xi \mathbf{I}_{N_{\mathrm{R}}}$ with $\beta = \left(\kappa_t^2 \frac{\rho}{N_{\mathrm{s}}} + \kappa_r^2 \frac{\rho}{N_{\mathrm{s}}} \mathbf{V}_{\Psi} \mathbf{V}_{\Psi}^H\right)$.

The optimisation problem, providing the precoders $\mathbf{V}_{\mathrm{RF}}, \mathbf{V}_{\mathrm{BB}}$, takes the form

$$\max_{\mathbf{V}_{\mathrm{RF}}^{\mathrm{opt}}, \mathbf{V}_{\mathrm{BB}}^{\mathrm{opt}}} \mathcal{I}(\mathbf{V}_{\Psi}) \tag{6.16a}$$

$$\text{s.t.} \quad \mathbf{V}_{\mathrm{RF}} \in \mathcal{V}_{\mathrm{RF}} \tag{6.16b}$$

$$\|\mathbf{V}_{\Psi}\|_F^2 = N_{\mathrm{s}}, \tag{6.16c}$$

with $\mathcal{V}_{\mathrm{RF}}$ denoting the set of $N_{\mathrm{T}} \times N_{\mathrm{T}}^{\mathrm{RF}}$ matrices that describe the feasible RF precoders having constant-magnitude entries. The problem (6.16a)–(6.16c) is very challenging due to the non-convexity of the constraint on $\mathbf{V}_{\mathrm{RF}}$ in equation (6.16b). Moreover, the solution to the optimal precoder even without the constraints is unknown[80]. In such case, we propose an approximation of equations (6.16a)–(6.16c) that will allow us to find near-optimal precoders, which can be implemented in practice.

First, we apply SVD to the channel matrix $\mathbf{H}$ in equation (6.15), i.e., we set $\mathbf{H} = \mathbf{S}\boldsymbol{\Sigma}\mathbf{U}^H$ so that (6.15) can be rewritten as

$$\mathcal{I}(\mathbf{V}_{\Psi}) = \log_2(|\mathbf{I}_{N_{\mathrm{s}}} + \left(\beta \mathbf{I}_{N_{\mathrm{s}}} + \xi \boldsymbol{\Sigma}^{-2}\right)^{-1} \mathbf{U}^H \mathbf{V}_{\Psi} \mathbf{V}_{\Psi}^H \mathbf{U}|). \tag{6.17}$$

Then, we partition the matrices $\boldsymbol{\Sigma}$ and $\mathbf{U}$ as $\boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_1 & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_2 \end{bmatrix}$ and $\mathbf{U} = [\mathbf{U}_1\, \mathbf{U}_2]$ and note that the optimal unconstrained unitary precoder for $\mathbf{H}$ is given by $\mathbf{V}_1$. Since $\mathbf{U}_1$ cannot be expressed by means of $\mathbf{V}_{\Psi}$, we are going to make $\mathbf{V}_{\Psi}$ approach the optimal $\mathbf{U}_1$ so that the mutual information corresponding to $\mathbf{U}_1$ and $\mathbf{V}_{\Psi}$ is approximately equivalent. Hence, the mutual information of $\mathbf{V}_{\Psi}$ can be approximated as

$$\mathcal{I}(\mathbf{V}_{\Psi}) \;\approx\; \log_2\left(\left|\mathbf{I}_{N_{\mathrm{s}}} + \mathbf{D}_1\right|\right) - \mathrm{tr}\left(\mathbf{I}_{N_{\mathrm{s}}} - \mathbf{U}_1^H \mathbf{V}_{\Psi} \mathbf{V}_{\Psi}^H \mathbf{U}_1\right). \tag{6.18}$$

where $\mathbf{D}_1 = \left(\beta^* \mathbf{I}_{N_{\mathrm{s}}} + \xi \boldsymbol{\Sigma}_1^{-2}\right)^{-1}$, with $\beta^* = \left(\kappa_t^2 \frac{\rho}{N_{\mathrm{s}}} + \kappa_r^2 \frac{\rho}{N_{\mathrm{s}}} \mathbf{I}\right)$. Here, we assume that $\mathbf{V}_{\Psi}\mathbf{V}_{\Psi}^H \approx \mathbf{I}$, which follows from the fact that we are trying to minimise the Euclidean distance between the optimal $\mathbf{U}_1$ and $\mathbf{V}_{\Psi}$.

Remark 6.2. *Note that the first term of* (6.18) *represents the mutual information due to the optimal precoder in the presence of RATHIs, while the second term describes the hybrid precoder impaired by the phase noise on the transmitter side.*

Now, the precoder design problem to find the optimal hybrid precoders that

maximises the mutual information can be written as

$$\min_{\mathbf{V}_{\mathrm{RF}}^{\mathrm{opt}},\mathbf{V}_{\mathrm{BB}}^{\mathrm{opt}}} \|\mathbf{U}_1 - \mathbf{V}_\Psi|\|_F \tag{6.19a}$$

$$\text{s.t.} \quad \mathbf{V}_{\mathrm{RF}} \in \mathcal{V}_{\mathrm{RF}} \tag{6.19b}$$

$$\|\mathbf{V}_\Psi\|_F^2 = N_{\mathrm{s}}, \tag{6.19c}$$

where equation (6.19a) follows since maximising tr $\left(\mathbf{U}_1^H \mathbf{V}_\Psi\right)$ is equivalent to minimising $\|\mathbf{U}_1 - \mathbf{V}_\Psi\|_F$. In other words, the objective is to find the projection of $\mathbf{U}_1 = \mathbf{V}^{\mathrm{opt}}$ on to the set of precoders having the form of $\mathbf{V}_\Psi = \mathbf{V}_{\mathrm{RF}}\mathbf{\Psi}\mathbf{V}_{\mathrm{BB}}$, with $\mathbf{V}_{\mathrm{RF}}$ belonging to the set of $\mathcal{V}_{\mathrm{RF}}$.

By exploiting the properties of the mmWave channel, we can obtain the near-optimal solution of problem described by equations (6.19a)–(6.19c). The specific properties to be employed are given as follows [28].

1. An orthonormal basis regarding the row space of the channel is formed by the columns of the unitary matrix $\mathbf{U}$.

2. A basis is formed for the same space by the linearly independent transmit antenna array response vectors $\mathbf{a}_{\mathrm{T}}(\phi, \theta)$ when $N_{\mathrm{cl}}N_{\mathrm{p}} \leq \min(N_{\mathrm{T}}, N_{\mathrm{R}})$[2].

3. Combining properties 1 and 2, the optimal unitary precoder $\mathbf{U}_1$ can be expressed as a linear combination of vectors $\mathbf{a}_{\mathrm{T}}(\phi_{\mathrm{T}}^{l,k}, \theta_{\mathrm{T}}^{l,k})$.

4. Given that $\mathbf{a}_{\mathrm{T}}(\phi_{\mathrm{T}}^{l,k}, \theta_{\mathrm{T}}^{l,k})$ are constant-magnitude phase-only vectors that can be applied at RF using analog phase shifters, the mmWave transmitter can employ $N_{\mathrm{T}}^{\mathrm{RF}}$ of the vectors $\mathbf{a}_{\mathrm{T}}(\phi_{\mathrm{T}}^{l,k}, \theta_{\mathrm{T}}^{l,k})$ and create arbitrary linear combinations of $\mathbf{a}_{\mathrm{T}}(\phi_{\mathrm{T}}^{l,k}, \theta_{\mathrm{T}}^{l,k})$ by using the digital precoder $\mathbf{V}_{\mathrm{BB}}$.

Accordingly, we can account for the restriction on the set $\mathcal{V}_{\mathrm{RF}}$ to $\mathbf{a}_{\mathrm{T}}(\phi_{\mathrm{T}}^{l,k}, \theta_{\mathrm{T}}^{l,k})$ and the precoder problem (6.19a)–(6.19c) can be written as

$$\min_{\mathbf{V}_{\mathrm{RF}}^{\mathrm{opt}},\mathbf{V}_{\mathrm{BB}}^{\mathrm{opt}}} \|\mathbf{U}_1 - \mathbf{V}_\Psi\|_F \tag{6.20a}$$

$$\text{s.t.} \quad \mathbf{V}_{\mathrm{RF}}^{(l)} \in \mathbf{a}_{\mathrm{T}}(\phi_{\mathrm{T}}^{l,k}, \theta_{\mathrm{T}}^{l,k}),\ \forall l, k \tag{6.20b}$$

$$\|\mathbf{V}_\Psi\|_F^2 = N_{\mathrm{s}}, \tag{6.20c}$$

[2]When $N_{\mathrm{cl}}N_{\mathrm{p}} \leq N_{\mathrm{T}}$, then the linear independence of $\mathbf{a}_{\mathrm{T}}(\phi, \theta)$ is guaranteed with probability one

Embedding $\mathbf{V}_{\mathrm{RF}}^{(l)}$ into the optimisation objective (6.20a), we have

$$\min_{\tilde{\mathbf{V}}_{\mathrm{BB}}} \|\mathbf{U}_1 - \mathbf{A}_{\mathrm{T}}\boldsymbol{\Psi}\tilde{\mathbf{V}}_{\mathrm{BB}}\|_F \tag{6.21a}$$

$$\text{s.t.} \|\mathrm{diag}\left(\tilde{\mathbf{V}}_{\mathrm{BB}}\tilde{\mathbf{V}}_{\mathrm{BB}}^H\right)\|_0 = N_{\mathrm{T}}^{\mathrm{RF}} \tag{6.21b}$$

$$\|\mathbf{A}_{\mathrm{T}}\tilde{\mathbf{V}}_{\mathrm{BB}}\|_F^2 = N_{\mathrm{s}}, \tag{6.21c}$$

where $\mathbf{A}_{\mathrm{T}} = \left[\mathbf{a}_{\mathrm{T}}(\phi_{\mathrm{T}}^{1,1}, \theta_{\mathrm{T}}^{1,1}), \ldots, \mathbf{a}_{\mathrm{T}}(\phi_{\mathrm{T}}^{N_{\mathrm{cl}},N_{\mathrm{p}}}, \theta_{\mathrm{T}}^{N_{\mathrm{cl}},N_{\mathrm{p}}})\right]$ is an $N_{\mathrm{T}} \times N_{\mathrm{cl}}N_{\mathrm{p}}$ matrix containing the array response vectors, while $\tilde{\mathbf{V}}_{\mathrm{BB}}$ is an $N_{\mathrm{cl}}N_{\mathrm{p}} \times N_{\mathrm{s}}$ matrix. The roles of $\mathbf{A}_{\mathrm{T}}$ and $\tilde{\mathbf{V}}_{\mathrm{BB}}$ in obtaining $\mathbf{V}_{\mathrm{RF}}^{\mathrm{opt}}$ and $\mathbf{V}_{\mathrm{BB}}^{\mathrm{opt}}$ are auxiliary as far as the sparsity constraint $\|\mathrm{diag}\left(\tilde{\mathbf{V}}_{\mathrm{BB}}\tilde{\mathbf{V}}_{\mathrm{BB}}^H\right)\|_0 = N_{\mathrm{T}}^{\mathrm{RF}}$ is concerned. This implies that $\tilde{\mathbf{V}}_{\mathrm{BB}}$ cannot have more than $N_{\mathrm{T}}^{\mathrm{RF}}$ non-zero rows.

The problem described by equations (6.21a)–(6.21c) is similar to the signal recovery approximation problems with multiple measurement vectors [116, 117, 154]. Hence, we apply the orthogonal matching pursuit method [3] to solve the problem as outlined in Algorithm 6.1. In particular, we begin by finding the vector $\mathbf{a}_{\mathrm{T}}(\phi_{\mathrm{T}}^{l,k}, \theta_{\mathrm{T}}^{l,k})$ along which the optimal precoder, $\mathbf{U}_1$, has its maximum projection in steps 3–6. Afterwards, the least-squares solution to $\mathbf{V}_{\mathrm{BB}}$ is obtained in step 7. In the next step, the algorithm finds the column for which the precoder has the largest projection. This column results from the removal of the chosen vector obtained from step 7. After $N_{\mathrm{T}}^{\mathrm{RF}}$ iterations, all the $N_{\mathrm{T}}^{\mathrm{RF}}$ beamforming vectors will have been selected, i.e.,the precoding matrix $\mathbf{V}_{\mathrm{RF}}$ will have been constructed. Also, the optimal baseband precoder $\mathbf{V}_{\mathrm{BB}}$ will have been found.

6.4.2 Hybrid combiner design

Given that the practical receivers will combine the received signals before detection, we engage in the design of the hybrid combiners $(\mathbf{W}_{\mathrm{RF}}, \mathbf{W}_{\mathrm{BB}})$ in this subsection to ensure that the MSE between the transmitted and received signals is minimised over a practical mmWave channel with RTHIs.

In this context, the optimisation problem providing the combiners $(\mathbf{W}_{\mathrm{RF}}, \mathbf{W}_{\mathrm{BB}})$

[3]The OMP method is most commonly in literature to design hybrid beamformers that achieve spectral efficiency for constrained systems comparable to that of the optimal unconstrained beamformer due to its simple implementation and low complexity than other methods such as the manifold optimisation [82] and the matrix decomposition [155].

takes the form

$$\min_{\mathbf{W}_{\mathrm{RF}}^{\mathrm{opt}},\mathbf{W}_{\mathrm{BB}}^{\mathrm{opt}}} \mathbb{E}\left[\|\mathbf{s}-\mathbf{W}_{\Omega}^{H}\mathbf{y}\|_2^2\right] \tag{6.22a}$$

$$\text{s.t.} \qquad \mathbf{W}_{\mathrm{RF}} \in \mathcal{W}_{\mathrm{RF}}, \tag{6.22b}$$

where $\mathbf{y}$ is the transmitted signal given in equation (6.2), $\mathbf{W}_{\Omega} = \mathbf{W}_{\mathrm{RF}}\boldsymbol{\Omega}\mathbf{W}_{\mathrm{BB}}$ and $\mathcal{W}_{\mathrm{RF}}$ denotes the set of $N_{\mathrm{R}} \times N_{\mathrm{R}}^{\mathrm{RF}}$ matrices that describe the feasible RF combiners and have constant-gain entries. Herein, we assume that the optimal precoders,

Algorithm 6.1: Orthogonal matching pursuit process to solve hybrid precoder problem.

Require $\mathbf{U}_1$, $\boldsymbol{\Psi}$
$\mathbf{V}_{\mathrm{RF}}$ = Empty Matrix
$\mathbf{V}_{\mathrm{res}} = \mathbf{V}_1$
for $i \leq N_{\mathrm{T}}^{\mathrm{RF}}$ **do**
  $\boldsymbol{\Pi} = \mathbf{A}_{\mathrm{T}}\boldsymbol{\Psi}\mathbf{V}_{\mathrm{res}}$
  $a = \max_{l=1,\ldots,N_{\mathrm{cl}}N_{\mathrm{p}}} \left(\boldsymbol{\Pi}\boldsymbol{\Pi}^{H}\right)_{l,l}$
  $\mathbf{V}_{\mathrm{RF}} = \left[\mathbf{V}_{\mathrm{RF}}|\mathbf{A}_{\mathrm{T}}^{(a)}\right]$
  $\mathbf{V}_{\mathrm{BB}} = \left(\mathbf{V}_{\mathrm{RF}}^{H}\mathbf{V}_{\mathrm{RF}}\right)^{-1}\mathbf{V}_{\mathrm{RF}}^{H}\mathbf{U}_1$
  $\mathbf{V}_{\mathrm{res}} = \frac{\mathbf{U}_1-\mathbf{V}_{\Psi}}{\|\mathbf{U}_1-\mathbf{V}_{\Psi}\|_F}$
end
$\mathbf{V}_{\mathrm{BB}} = \sqrt{N_{\mathrm{s}}}\frac{\mathbf{V}_{\mathrm{BB}}}{\|\mathbf{V}_{\mathrm{RF}}\mathbf{V}_{\mathrm{BB}}\|_F}$
Return $\mathbf{V}_{\mathrm{RF}}$,$\mathbf{V}_{\mathrm{BB}}$

($\mathbf{V}_{\mathrm{RF}}^{\mathrm{opt}}$, $\mathbf{V}_{\mathrm{BB}}^{\mathrm{opt}}$), which have been determined by the transmitter, are fixed. Analogous to the precoder design, no general solutions to equation (6.22a) are known, due to the non-convexity of the constraint $\mathbf{W}_{\mathrm{RF}} \in \mathcal{W}_{\mathrm{RF}}$.[4] Hence, we provide a transformation of the problem (4.51a)–(4.47b) which will allow us to find the optimal hybrid combiners that minimise the MSE between the transmitted and received signals. First, expanding the objective function of equation (6.22a) we have

$$\mathbb{E}\left[\|\mathbf{s}-\mathbf{W}_{\Omega}^{H}\mathbf{y}\|_2^2\right] = \mathrm{tr}\left(\mathbb{E}\left[\mathbf{s}\mathbf{s}^{H}\right]\right) - 2\mathbb{R}\{\mathrm{tr}\left(\mathbb{E}\left[\mathbf{s}\mathbf{y}^{H}\right]\mathbf{W}_{\Omega}\right)\} + \mathrm{tr}\left(\mathbf{W}_{\Omega}^{H}\mathbb{E}\left[\mathbf{y}\mathbf{y}^{H}\right]\mathbf{W}_{\Omega}\right). \tag{6.23}$$

[4] Interestingly, an exact solution to the MSE optimisation problem in equation (6.22a) without hardware impairments would have been possible with the implementation of the matrix inversion lemma [28, 156].

The optimisation problem described by equation (4.47a)–(4.47b) is conditioned only on $\mathbf{W}_{\mathrm{RF}}$ and $\mathbf{W}_{\mathrm{BB}}$. This implies that the addition of any constant which is independent of $\mathbf{W}_{\mathrm{RF}}$ and $\mathbf{W}_{\mathrm{BB}}$ to the objective function would not change the solution of equations (4.47a)–(4.47b). Accordingly, equation (6.23) can be rewritten as

$$\begin{aligned}\mathbb{E}\left[\|\mathbf{s}-\mathbf{W}_{\Omega}^{H}\mathbf{y}\|_2^2\right] &= \mathrm{tr}\left(\mathbf{W}_{M}^{H}\mathbb{E}\left[\mathbf{y}\mathbf{y}^{H}\right]\mathbf{W}_{M}\right)-\mathrm{tr}\left(\mathbb{E}\left[\mathbf{s}\mathbf{s}^{H}\right]\right)+\mathrm{tr}\left(\mathbb{E}\left[\mathbf{s}\mathbf{s}^{H}\right]\right)\\ &- 2\mathbb{R}\{\mathrm{tr}(\mathbb{E}\left[\mathbf{s}\mathbf{y}^{H}\right]\mathbf{W}_{\Omega})\}+\mathrm{tr}(\mathbf{W}_{\Omega}^{H}\mathbb{E}\left[\mathbf{y}\mathbf{y}^{H}\right]\mathbf{W}_{\Omega}). \qquad (6.24)\end{aligned}$$

Here, we have introduced the term $\mathrm{tr}\left(\mathbf{W}_{M}^{H}\mathbb{E}\left[\mathbf{y}\mathbf{y}^{H}\right]\mathbf{W}_{M}\right)-\mathrm{tr}\left(\mathbb{E}\left[\mathbf{s}\mathbf{s}^{H}\right]\right)$, where $\mathbf{W}_{M}$ is the common conventional solution to the unconstrained minimum MSE problems given by

$$\begin{aligned}\mathbf{W}_{M}^{H} &= \mathbb{E}\left[\mathbf{s}\mathbf{y}^{H}\right]\mathbb{E}\left[\mathbf{y}\mathbf{y}^{H}\right]^{-1}\\ &= \frac{1}{\sqrt{\rho}}\left(\mathbf{V}_{\Psi}^{H}\mathbf{H}^{H}\mathbf{H}\mathbf{V}_{\Psi}+\frac{\sigma^2 N_{\mathrm{s}}}{\rho}\mathbf{I}_{N_{\mathrm{s}}}\right)^{-1}\mathbf{V}_{\Psi}^{H}\mathbf{H}^{H}. \qquad (6.25)\end{aligned}$$

Substituting $\mathbb{E}\left[\mathbf{s}\mathbf{y}^{H}\right]=\mathbf{W}_{M}^{H}\mathbb{E}\left[\mathbf{y}\mathbf{y}^{H}\right]$ into equation (6.24), we have

$$\begin{aligned}\mathbb{E}\left[\|\mathbf{s}-\mathbf{W}_{\Omega}^{H}\mathbf{y}\|_2^2\right] &= \mathrm{tr}\left(\mathbf{W}_{M}^{H}\mathbb{E}\left[\mathbf{y}\mathbf{y}^{H}\right]\mathbf{W}_{M}\right)-2\mathbb{R}\{\mathrm{tr}\left(\mathbb{E}\left[\mathbf{W}_{M}^{H}\mathbb{E}\left[\mathbf{y}\mathbf{y}^{H}\right]\mathbf{W}_{\Omega}\right]\right)\}\\ &\quad+\mathrm{tr}\left(\mathbf{W}_{\Omega}^{H}\mathbb{E}\left[\mathbf{y}\mathbf{y}^{H}\right]\mathbf{W}_{\Omega}\right)\\ &= \mathrm{tr}\left((\mathbf{W}_{M}^{H}-\mathbf{W}_{\Omega}^{H})\,\mathbb{E}\left[\mathbf{y}\mathbf{y}^{H}\right](\mathbf{W}_{M}^{H}-\mathbf{W}_{\Omega}^{H})^{H}\right) \qquad (6.26)\\ &\overset{(a)}{=} \|\mathbb{E}\left[\mathbf{y}\mathbf{y}^{H}\right]^{1/2}(\mathbf{W}_{M}-\mathbf{W}_{\Omega})\|_{F},\end{aligned}$$

where (a) follows from $\|A\|_F=\sqrt{\mathrm{tr}(AA^H)}$. Due to equation (6.26), the optimisation problem to find the optimal hybrid combiners becomes

$$\min_{\mathbf{W}_{\mathrm{RF}}^{\mathrm{opt}},\mathbf{W}_{\mathrm{BB}}^{\mathrm{opt}}} \|\mathbb{E}\left[\mathbf{y}\mathbf{y}^{H}\right]^{1/2}(\mathbf{W}_{M}-\mathbf{W}_{\Omega})\|_{F} \qquad (6.27a)$$

$$\text{s.t.} \qquad \mathbf{W}_{\mathrm{RF}}\in\mathcal{W}_{\mathrm{RF}},, \qquad (6.27b)$$

In other words, we aim to find the projection of the unconstrained combiner $\mathbf{W}_{M}$, weighted by $\mathbb{E}\left[\mathbf{y}\mathbf{y}^{H}\right]$, on to the set of combiners having the form $\mathbf{W}_{\Omega}=\mathbf{W}_{\mathrm{RF}}\boldsymbol{\Omega}\mathbf{W}_{\mathrm{BB}}$. Comparable to the design of the optimal precoders in section 6.4.1, it is algorithmically intractable to find this projection in closed form due to the non-convexity of the set $\mathcal{W}_{\mathrm{RF}}$. Accordingly, we exploit the aforementioned properties of the mmWave channel to result in a near-optimal solution to (6.27a).

Specifically, an orthonormal basis is formed for the columns of $\mathbf{W}_{\mathrm{RF}}$ by the linearly independent array response vectors, $\mathbf{a}_{\mathrm{R}}(\phi_{\mathrm{R}}^{l,k}, \theta_{\mathrm{R}}^{l,k})$, so that the receiver utilises $N_{\mathrm{R}}^{\mathrm{RF}}$ of the vectors $\mathbf{a}_{\mathrm{R}}(\phi_{\mathrm{R}}^{l,k}, \theta_{\mathrm{R}}^{l,k})$ at the RF using analog phase shifters. For this reason, we can rewrite the optimisation problem (6.27a)–(6.27b) by restricting the columns of $\mathbf{W}_{\mathrm{RF}}$ to have the form of $\mathbf{a}_{\mathrm{R}}(\phi_{\mathrm{R}}^{l,k}, \theta_{\mathrm{R}}^{l,k})$, as follows:

$$\min_{\mathbf{W}_{\mathrm{RF}}^{\mathrm{opt}}, \mathbf{W}_{\mathrm{BB}}^{\mathrm{opt}}} \|\mathbb{E}\left[\mathbf{y}\mathbf{y}^H\right]^{1/2} (\mathbf{W}_M - \mathbf{W}_\Omega)\|_F \tag{6.28a}$$

$$\text{s.t.} \qquad \mathbf{W}_{\mathrm{RF}}^{(l)} \in \mathbf{a}_{\mathrm{R}}(\phi_{\mathrm{R}}^{l,k}, \theta_{\mathrm{R}}^{l,k}),\ \forall l, k. \tag{6.28b}$$

Herein, the task is to find the optimal combiner in terms of the optimal basis formed by $\mathbf{a}_{\mathrm{R}}(\phi_{\mathrm{R}}^{l,k}, \theta_{\mathrm{R}}^{l,k})$. Moreover, inserting $\mathbf{W}_{\mathrm{RF}}^{(l)}$ into the objective function (6.28a) results in equations (6.29a)–(6.29b) which use $\mathbf{A}_{\mathrm{R}}$ and $\tilde{\mathbf{W}}_{\mathrm{BB}}$ as auxiliary matrices to obtain $\mathbf{W}_{\mathrm{BB}}^{\mathrm{opt}}$:

$$\min_{\tilde{\mathbf{W}}_{\mathrm{BB}}} \|\mathbb{E}\left[\mathbf{y}\mathbf{y}^H\right]^{\frac{1}{2}} \mathbf{W}_M - \mathbb{E}\left[\mathbf{y}\mathbf{y}^H\right]^{\frac{1}{2}} \mathbf{A}_{\mathrm{R}}\, \Omega\, \tilde{\mathbf{W}}_{\mathrm{BB}}\|_F \tag{6.29a}$$

$$\text{s.t.} \|\mathrm{diag}\left(\tilde{\mathbf{W}}_{\mathrm{BB}} \tilde{\mathbf{W}}_{\mathrm{BB}}^H\right)\|_0 = N_{\mathrm{R}}^{\mathrm{RF}}, \tag{6.29b}$$

where $\mathbf{A}_{\mathrm{R}} = \left[\mathbf{a}_{\mathrm{R}}(\phi_{\mathrm{R}}^{1,1}, \theta_{\mathrm{R}}^{1,1}), \ldots, \mathbf{a}_{\mathrm{R}}(\phi_{\mathrm{R}}^{N_{\mathrm{cl}}, N_{\mathrm{p}}}, \theta_{\mathrm{R}}^{N_{\mathrm{cl}}, N_{\mathrm{p}}})\right]$ is an $N_{\mathrm{R}} \times N_{\mathrm{cl}} N_{\mathrm{p}}$ matrix containing the array response vectors, while $\tilde{\mathbf{W}}_B$ is an $N_{\mathrm{cl}} N_{\mathrm{p}} \times N_{\mathrm{s}}$ matrix.

Algorithm 6.2: OMP-based algorithm for hybrid combiners.

Require $\mathbf{W}_M$, $\mathbf{\Omega}$
$\mathbf{W}_{\mathrm{RF}}$ = Empty Matrix
$\mathbf{W}_{\mathrm{res}} = \mathbf{W}_M$
for $i \leq N_{\mathrm{R}}^{\mathrm{RF}}$ **do**
  $\mathbf{\Pi} = \mathbf{A}_r^H\, \mathbf{\Omega}\, \mathbb{E}\left[\mathbf{y}\mathbf{y}^H\right] \mathbf{W}_{\mathrm{res}}$
  $a = \max_{l=1,\ldots,N_{\mathrm{cl}} N_{\mathrm{p}}} \left(\mathbf{\Pi}\mathbf{\Pi}^H\right)_{l,l}$
  $\mathbf{W}_{\mathrm{RF}} = \left[\mathbf{W}_{\mathrm{RF}} | \mathbf{A}_{\mathrm{R}}^{(a)}\right]$
  $\mathbf{W}_{\mathrm{BB}} = \left(\mathbf{W}_{\mathrm{RF}}^H \mathbb{E}\left[\mathbf{y}\mathbf{y}^H\right] \mathbf{W}_{\mathrm{RF}}\right)^{-1} \mathbf{W}_{\mathrm{RF}}^H \mathbb{E}\left[\mathbf{y}\mathbf{y}^H\right] \mathbf{W}_M$
  $\mathbf{W}_{\mathrm{res}} = \frac{\mathbf{W}_M - \mathbf{W}_\Omega}{\|\mathbf{W}_M - \mathbf{W}_\Omega\|_F}$
end
Return $\mathbf{W}_{\mathrm{RF}}$,$\mathbf{W}_{\mathrm{BB}}$

The optimisation problem described by equations (6.29b)–(6.29b) can also be solved by using the orthogonal matching pursuit method as outlined in Algorithm 6.2.

Remark 6.3. *Note that the design of the combiners relies on the decoupling of the precoding and combining processes. Despite the fact that the decoupling process decreases the complexity involved in the design of hybrid precoders and combiners in mmWave systems with RTHIs, the process simplifies the transceiver design. Hence, care must be taken to avoid an oversimplification of the process which can lead to received power losses. In particular, to avoid this, it becomes pertinent to start the design with the more constrained side of the mmWave system, i.e., either the transmitter or receiver side with the lowest number of* RF *chains.*

6.5 Numerical Results

Our simulations verify the analytical transceiver design results provided by Algorithms 6.1 and 6.2 and illustrate the impact of the various hardware impairments on the spectral efficiency of the mmWave system. We consider a simulation set-up where the transmitter communicates by means of a single stream to the receiver with 8 clusters and 10 rays per cluster. Throughout the simulations, the azimuth and elevation AoAs and AoDs are Laplacian-distributed with an angular spread of 7.5° and the transmit power ρ is dependent on the signal to

Table 6.1: Parameter settings for simulation.

Notation	**Parameter**	**Value**
N_{T}	Number of transmit antennas	64
N_{R}	Number of receive antennas	16
f_c	Carrier frequency	30 GHz
c_n	Oscillator constant	10^{-17}
σ^2	Thermal noise variance	-174 dbm/Hz
N_{s}	Number of transmit streams	1
$N_{\mathrm{T}}^{\mathrm{RF}}$	Number of transmit RF chains	4
$N_{\mathrm{R}}^{\mathrm{RF}}$	Number of receive RF chains	4
BW	Bandwidth	100 MHz
κ_{R}^2	RATHI constant at the receiver	0.0156
κ_{T}^2	RATHI constant at the transmitter	0.0156

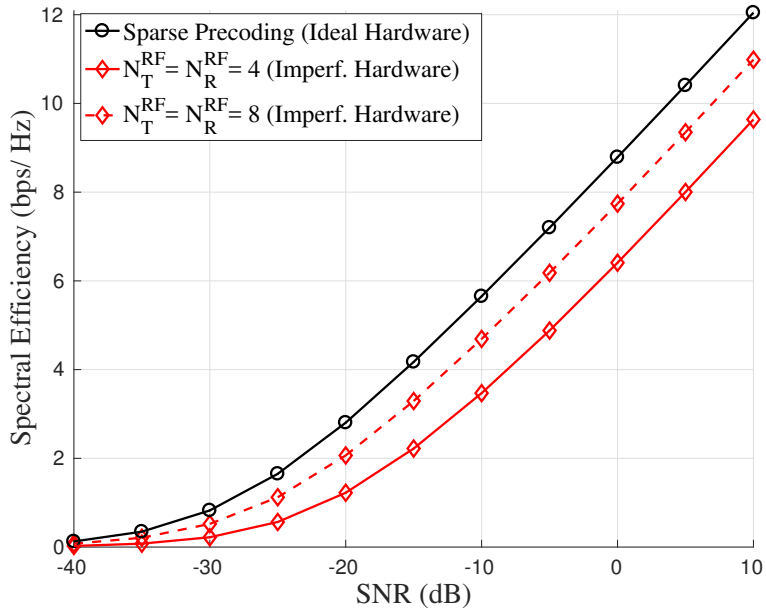


Figure 6.2: Spectral efficiency versus the SNR achieved by a 16×64 mmWave MIMO system showing the impact of PN with different number of RF chains.

noise ratio i.e., SNR $= \frac{\rho}{\sigma_z^2}$. We assume equal power allocation in all precoding and combining solutions. The PN is simulated as a discrete Wiener process, and the ATN is Gaussian-distributed. A set of the parameters, inspired by related studies on non-ideal hardware and mmWave communications [28, 135], with their corresponding values are presented in Table 6.1. The objective of this section is to compare the results obtained by ideal and non-ideal hardware (each separate hardware impairment) in order to understand which impairments have a greater impact and what conditions can provide improved system performance.

6.5.1 Multiplicative impairments (PN)

In Figs. 6.2 and 6.3, we investigate only the impact of PN on the spectral efficiency achieved by means of hybrid precoding and combining on a 16×64 mmWave MIMO system. As stated earlier in section 2.3.2, the attractiveness of hybrid precoding in mmWave communications comes from the fact that it provides a balance between hardware complexity and system performance. Given that in hybrid precoding the number of RF chains can vary from one to the maximum number of antennas (as in full digital precoding), in Fig. 6.2 we investigate the impact of PN on the spectral efficiency when the number of RF chains is increased.

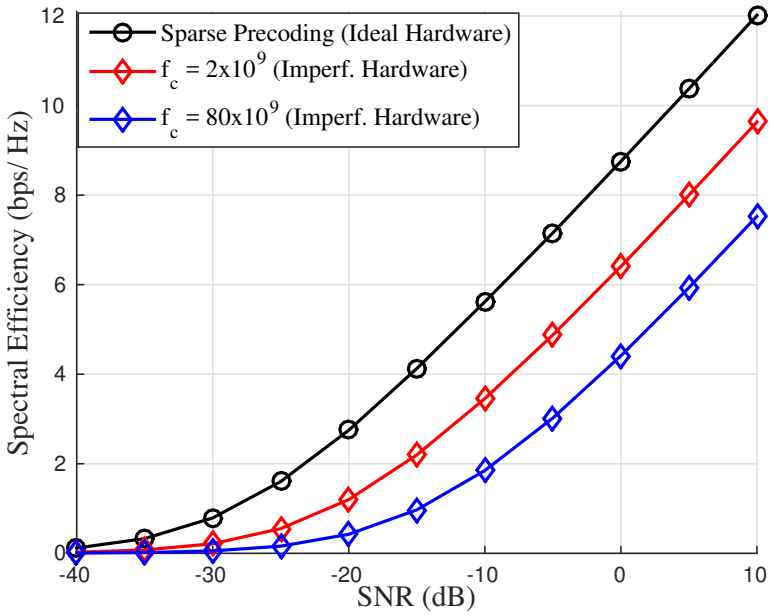


Figure 6.3: Spectral Efficiency versus the SNR achieved by a 16×64 mmWave MIMO system showing the impact of PN under different mmWave carrier frequencies with $N_{\mathrm{T}}^{\mathrm{RF}} = N_{\mathrm{R}}^{\mathrm{RF}} = 4$ RF chains.

From this figure, it can be seen that the PN degrades the spectral efficiency much more when 4 RF chains are used in spatial hybrid precoding than when 8 RF chains are employed in the design. This can be explained by the fact that when the LOs share a common oscillator, they experience the same PN, thus increasing the number of RF chains moves the SE closer to that obtained in the fully digital case. The result is striking because while the RF chains could be reduced to decrease the complexity and communication costs, especially in mmWave systems with large antenna arrays, we observe a degradation on the spectral efficiency performance due to PN. Hence, this implies a remarkable design consideration. However, we note that the results may differ when the LOs do not share a common oscillator and have different PNs.

Fig. 6.3 provides a comparison of the spectral efficiency versus the SNR for different carrier frequencies. It can be observed from the figure that the degradation of the spectral efficiency increases due to the PN as the communication moves to higher mmWave frequencies. This result is important and implies that in practical scenarios the PN must be considered during the choice of the mmWave band for the achievement of communication.

6.5.2 RATHIs

Next, in Figs 6.4 and 6.5, we investigate how the RATHIs affect the spectral efficiency. In other words, we assume the absence of phase and amplified thermal noises. Specifically, in Fig. 6.4, we present the spectral efficiency against SNR under varying values of transmit and receive RATHI constants (κ_{T}^2 and κ_{R}^2), respectively. These values are not arbitrary since they have been adopted from system measurements in [135]. It can be observed that the impact of the RATHIs is negligible at low SNR values, under 0 dB, as indicated by the saturation of the spectral efficiency performance with respect to the ideal case. However, for SNR values above 0 dB, the degradation of spectral efficiency grows with increasing values of κ_{T}^2 and κ_{T}^2. This result is particularly useful for mmWave systems because the mmWave transmissions are characterised by low signal-to-noise ratios, implying that the distortion noises will have little impact on the system's performance. Another important observation is depicted in Fig. 6.5, which shows that the transmit and receive distortion impairments contribute the same amount of degradation to the spectral efficiency. As illustrated in the figure, when we set $\kappa_{\mathrm{R}}^2 = 0$ and $\kappa_{\mathrm{T}}^2 = 0.05$ or $\kappa_{\mathrm{T}}^2 = 0$ and $\kappa_{\mathrm{R}}^2 = 0.05$, the spectral efficiency does not change. This observation is quite interesting and useful. It suggests that in the case of communication at the high SNR regime, in order to achieve a desired spectral efficiency, a design scheme can be chosen by keeping κ_{R}^2 constant and varying κ_{T}^2, which determines the quality of transmitting hardware i.e, how prone the transmitter is to impairments.

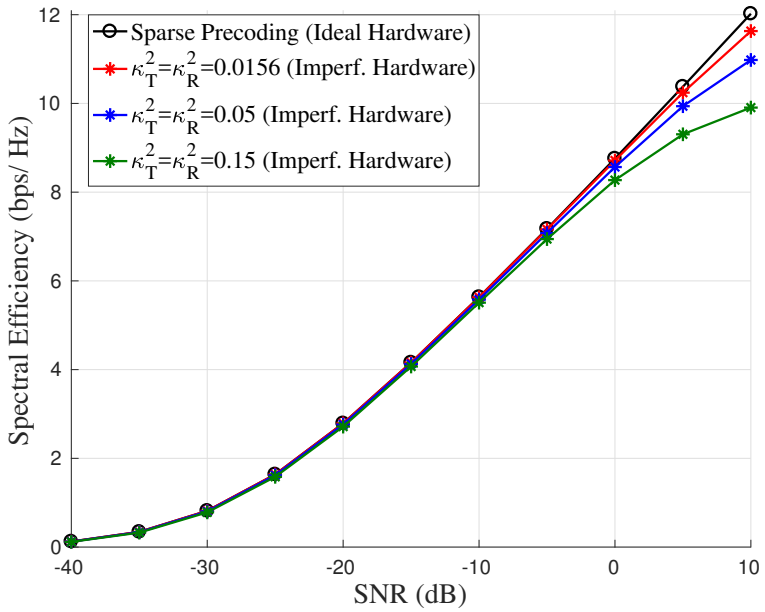


Figure 6.4: Spectral efficiency versus the SNR achieved in a 16×64 mmWave MIMO system showing the impact of the RATHIs.

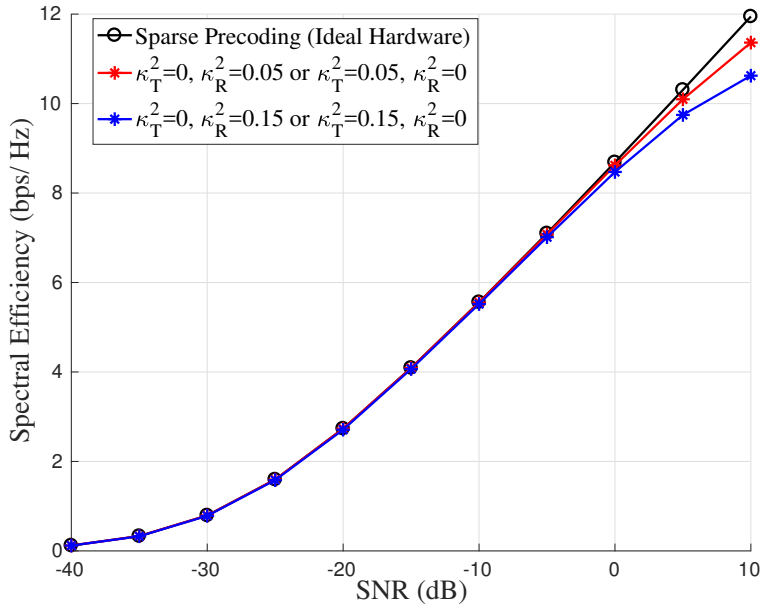


Figure 6.5: Spectral efficiency versus the SNR achieved in a 16×64 mmWave MIMO system showing the impact of the RATHIs.

6.5.3 ATN

In order to demonstrate the effect of ATN on the system's performance, we assume that the rest of the impairments, i.e., the additive distortion and phase noises, are

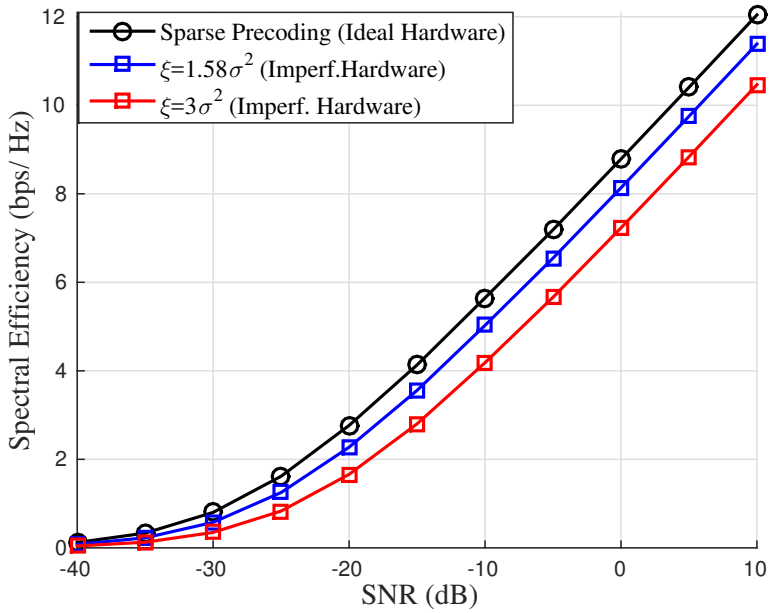


Figure 6.6: Spectral efficiency versus the SNR achieved in a 16×64 mmWave MIMO system showing the impact of ATN.

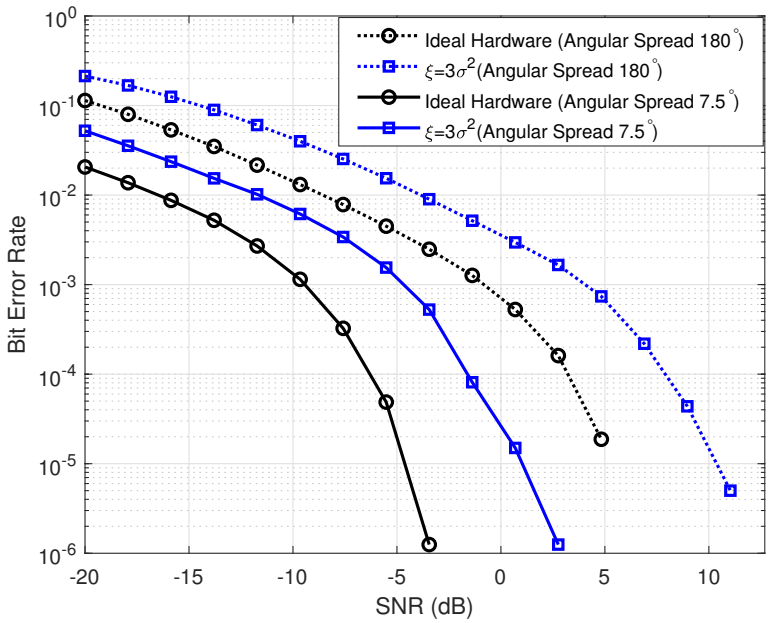


Figure 6.7: BER performance achieved in a 16×64 mmWave MIMO system showing the impact of ATN under 4 QAM signalling.

considered to have no effect. Herein, we plot the spectral efficiency and bit error rate versus the SNR for different values of ξ taken from circuit measurements

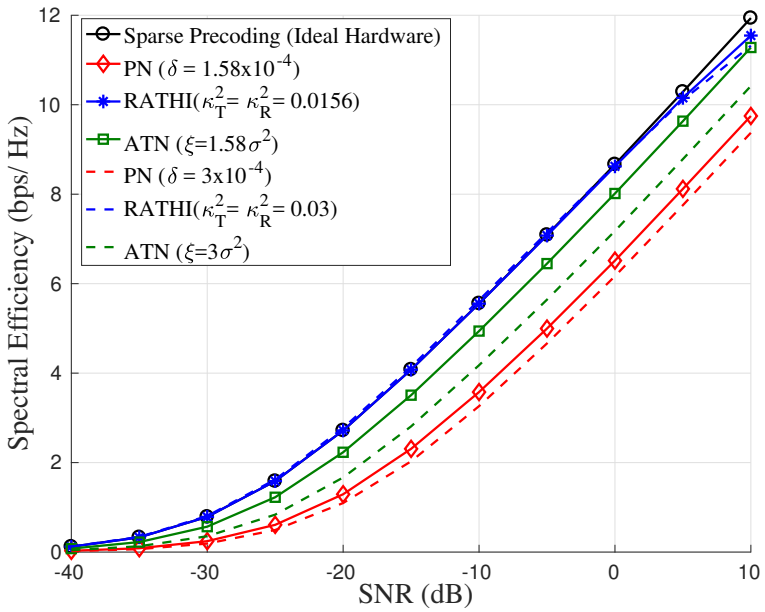


Figure 6.8: Comparison of the spectral efficiency versus SNR achieved in a 16×64 mmWave MIMO system showing the impact of hardware imperfections (PN, RATHI and ATN).

[135]. Specifically, in Fig. 6.6, it is shown that there is a steady degradation in spectral efficiency performance across all SNR values and that this degradation worsens with an increment in the value of ξ.

In Fig. 6.7, the variance of the ATN is set to $\xi = 3\sigma^2$. It can be observed that the error rate performance under the ideal hardware assumption decreases with the improved channel conditions. In other words, the wider the angular spread of the angles (AoAs and AoDs) used in transmission, the higher the probability of error occurrence in the received signal. To quantify, to achieve a BER of 10^{-2}, transmission with an angular spread of 7.5° achieves an approximate 6-dB gain when compared with the error rate performance obtained by transmitting with an angular spread of 180°. Nevertheless, it can be noted that the amplified thermal noise still degrades the performance at a steady rate. These results imply that the amplified thermal noise must be considered in the planning of mmWave communication systems.

Next, Figs 6.8 and 6.9 present a comparison of the spectral efficiency by using ideal hardware and using imperfect hardware. In Fig. 6.8, the imperfect hardware parameters are set to $\kappa_T^2 = \kappa_R^2 = 0.0156$, $\xi = 1.58\sigma^2$, $\delta = 1.58 \times 10^{-4}$ and $\kappa_T^2 = \kappa_R^2 = 0.03$, $\xi = 3\sigma^2$, $\delta = 3 \times 10^{-4}$. These baseline hardware

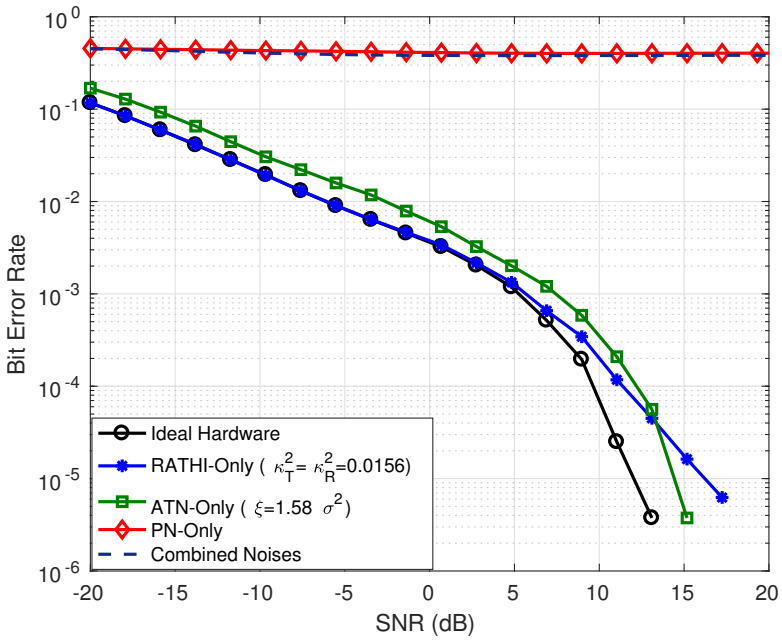


Figure 6.9: Comparison of the BER performance versus SNR achieved in a 16×64 mmWave MIMO system showing the impact of hardware imperfections (PN, RATHI and ATN) under 4 QAM signalling.

imperfection values differ by the same ratio; in other words, the two values of PN, ATN and RATHI have the same ratio of $\frac{13}{25}$. It can be observed that the PN and the ATN increase monotonically with the SNR. Evidently, multiplicative PN has the greatest impact, causing the most degradation in spectral efficiency when compared with the performance using ideal hardware. Notably, however, when comparing the degradation between each separate impairment for the given specific ratio, ATN performs the worst. To quantify, at 5 dB, we observe a 2.45%, 10.22% and 0.99% decrease in spectral efficiency performance from the impact of the PN, the ATN and the RATHI, respectively. Thus, the impact of ATN is increasingly critical for higher values of ξ.

Finally, from Fig. 6.9, which illustrates both the bit error rate performance of each separate impairment and the combined effects of all noises together, it can be observed that RATHI have the same error performance as the ideal hardware in low SNR regions and only affect the system in high SNR regimes. This result validates the findings from Figs 6.4 and 6.5 which suggest that RATHIs can be neglected in mmWave system design with hybrid precoding as their effect is negligible. For the ATN-only effect, it can be observed that an increase in

the number of errors is steady at all SNR values. It is apparent from the error floor resulting from the effect of PN only that this multiplicative noise is highly critical in mmWave systems and must not be neglected in the system design. In addition, the combined effect of all noises on the error performance also results in an error floor indicating the dominance of the PN. This result further highlights the significance of PN and provides an opportunity to be explored for future work, i.e., the design of the mmWave systems that are robust to multiplicative noise in the transceiver hardware.

6.6 Conclusion

Hybrid precoding beamforming solutions are prominent in mmWave MIMO communications as they provide a balance between complexity and spectral efficiency of the systems. Even though hardware imperfections are residual in the components that make up a communication system, their effects have not been taken into account as most prior studies have assumed perfect hardware while studying the performance of mmWave networks. The objective of this chapter was to shed light on the impact of transceiver hardware impairments on mmWave massive MIMO systems with hybrid precoding. Specifically, we designed hybrid precoding and combining solutions which account for the three major hardware impairments: the ATN, the RATHIs and the multiplicative PN. Simulation results which considered the impact of each impairment separately on the spectral efficiency performance of the system show that the multiplicative PN causes the highest degradation to the system's performance. The RATHIs degrade the system only at the high SNR regime, and the ATN causes a steady degradation to the spectral efficiency. After examining how varying the number of RF chains affects the impact of RATHIs, we noted that increasing the number of RF chains reduces the impact of PN, which is the most critical of the three hardware impairments on mmWave MIMO systems. This can be considered as a design choice to improve the spectral efficiency performance of mmWave MIMO systems with hybrid precoding in the presence of hardware transceiver impairments.

Chapter 7

Conclusions

7.1 Summary of Contributions

Mobile communication systems are experiencing a steady period of exponential growth in traffic, with smart wireless devices becoming ubiquitous, and the rapid proliferation of social media. This fast-paced growth is straining the existing digital wireless networks, and is pushing for new technologies that can support the expectations of ever-increasing data rates.

While improvements in spectral efficiencies can be achieved using physical layer techniques such as channel coding, MIMO and network densification, the spectrum crunch in today's cellular networks is a major obstacle to significant capacity increase and flexibility in spectrum management and usage, leading to a reduction in the scope of achievable improvements. Thus, it is imperative to shift the communication paradigm by investigating bands of the spectrum that have not been previously utilised for mobile communications.

To this end, the research in this thesis has addressed this pressing issue by studying mobile communication at millimeter wave (mmWave) frequencies in order to exploit the large amounts of unused bandwidth available within these bands. We began by investigating the peculiarities of communication at higher bands of the radio spectrum, and provided a tractable network architecture for the deployment of mmWave systems in Chapter 3. Specifically, we analysed the downlink performance of a multi-user mmWave CRAN system which employed hybrid beamforming in both fronthaul and access links. Stochastic geometry was used to characterise the spatially distributed RRHs and users in the system. We derived analytical expressions of the outage probability, average data latency and

throughput for two scenarios – when a typical user is served by the best RRH in the network, and when the nearest RRH to the typical user is employed in transmission. Additionally, we considered both noise-limited and interference-limited settings to model different practical deployments of mmWave CRAN systems. Results from that chapter have shown that although there exists a trade-off between density of RRHs and inter-cluster interference, the deployment of larger antenna arrays can compensate for the degradation of communication in terms of delivery latencies, throughput and outage probabilites with higher RRH deployment. This information could be used by engineers to make intelligent decisions about maintaining a high-performance network.

In Chapter 4, to show that mmWave systems have the potential to provide the device connectivity requirement of future generation communication networks, we studied the energy-efficient hybrid transceiver design of mmWave interference channels involving IoT devices. The system under consideration accounted for the limited dynamic range at the transmitters and receivers. Since the joint hybrid transceiver design problems are known to be non-convex, we transformed them into their corresponding univariate maximisation problems and proposed WMMSE alternating algorithms that are guaranteed to converge. Then, using the concept of orthogonal matching pursuit, we proposed a hybrid transceiver design algorithm that maximises the energy efficiency (EE) of the IoT network. Furthermore, the design was extended to account for imperfect CSI scenarios. Simulation results showed that the EE performance of the proposed algorithms is comparable to results obtained by unconstrained digital beamforming with lower complexity, indicating that mmWave systems can support IoT communication for both perfect and imperfect CSI considerations, and give a more environmentally friendly, energy-efficient system with lower complexity and high data rates.

Moreover, in consideration of real-life practical constraints such as transmit power restrictions, hardware imperfections and limited feedback overhead, we considered the impact of hardware impairments on the performance of mmWave systems in Chapter 6, and investigated the rate-splitting (RS) transmission strategy for multi-user mmWave systems with hybrid beamforming under imperfect CSI in Chapter 5. The goal of Chapter 5 was to alleviate the adverse effects of multi-user interference (MUI) resulting from imperfect CSI. The inherent sparsity of mmWave channels was exploited to estimate the dominant paths of the channel for each user. Based on this transmission strategy, the transmitted signal was divided into a common message and a private message, then the transmit power

was dynamically split between the private and common parts of the transmitted signal. In addition, we proposed an alternating maximization algorithm to provide an optimal RS precoder that guarantees interference-free transmission for the common message. Results show that the residual MUI in the system resulting from channel estimation errors is tackled by RS as evidenced by the improvement in sum-rate performance over conventional linear transmission techniques. In Chapter 6, we developed a framework for modelling three main categories of hardware imperfections, namely multiplicative phase noise, additive distortion noise and amplified thermal noise, and analysed the impact of each noise category on the achievable spectral efficiency of the system. It is evident from our analysis that phase noise is the most critical of the three noise categories, and it was found that an increase in the number of RF chains can reduce some of the detrimental effects of phase noise on the system. This observation provides information that can be used in the design of hybrid beamforming solutions which are prominent in mmWave communication systems.

Overall, across Chapters 3 to 6 of this thesis, we have contributed an improvement to the understanding of how the mmWave spectrum can be used to achieve mobile communications, and we have shown that it is possible to deploy environmentally friendly and energy-efficient mmWave systems with low complexity that provide very low latencies and ultra-high data rates. In other words, mmWave communications will be used to satisfy the requirements of the next-generation communication networks.

7.2 Limitations and Scope for Further Work

The work presented in this thesis provides fundamental insights to the workings of mmWave communications and brings to light several interesting areas for future research. In this section, we present some of the possible extensions to this work.

7.2.1 Fronthaul interface analysis

As mentioned earlier in section 3.2, the interface of the fronthaul is constrained by strict requirements of the CPRI standards which split functionally between the BBU and RRHs in the network. While the analysis presented in Chapter 3 gives an overview of the functionality split, it focuses on the access link transmissions

between the RRHs and UEs. Therefore, specific studies providing performance analysis that looks into the different functional splits of the fronthaul interface will be of great interest. More recently, a new radio interface specification for 5G has been released (eCPRI [157]) which allows a hardware and a software component for the BBU, implying that physical-layer functionality can be split between the BBU and RRHs, reducing the complexity of the RRHs and decreasing the data rate demands in the fronthaul interface. Such studies would be worthwhile to provide further insights into mmWave CRAN systems.

7.2.2 Hybrid beamforming architectures

- The hybrid beamforming framework employed in this thesis (Chapters 4–6) focused on the phase shifters' antenna configuration. This framework was chosen as it is well established in literature for advanced analog beamforming. However, there are other recently-proposed analog strategies that can be applied to hybrid beamforming, as mentioned in Chapter 2. Therefore, the analysis in Chapter 5, and the hybrid beamforming solutions in Chapters 4 and 6, can be extended for different antenna configurations, i.e., switches and lenses. In addition, comparing these hybrid antenna configurations would be worthwhile in finding relative advantages of different hybrid beamforming solutions.

- Another restriction of the phase shifters hybrid beamforming framework utilised in Chapters 4–6 is the assumption of phase-only control in the analog domain. This assumption while valid for downlink scenarios when the number of antennas at the receiver-end is small compared to the number of transmit antennas, neglects amplitude-control which is necessary for the management of side lobes and out-of-band-emissions. Therefore channel - reconstruction-based hybrid beamforming frameworks which consider both amplitude and phase control of the individual antennas in the RF domain will be of great interest and provide more generalised results, complementing the studies of Chapters 4–6.

- A related area worth exploring is the use of low-precision receivers as a solution to the power consumption bottleneck of the MIMO operation in mmWave systems. These low-precision receivers can also be employed in conjunction with a hybrid beamforming framework to further minimise the

power costs. Hence, the work of Chapters 4–6 can be extended using low-bit ADC receivers.

7.2.3 Imperfect CSI considerations

- The mmWave CRAN study in Chapter 3 provides analytical expressions to measure the outage probability, throughput and average latency of the system. However, the derivations were based on the idealistic assumption of the availability of perfect channel state at the BBU. Incorporating channel uncertainty into the analysis, where imperfect CSI is available at the BBU, will be of great interest and wider application.

- The EE study in Chapter 4 shows that energy-efficient hybrid beamformers can be robustly designed for mmWave systems with low complexity by incorporating channel uncertainty into the hybrid beamforming design, and the derivations were based on the assumption of the stochastic error model. In practice, however, the error model will vary based on the error type. For instance, if quantisation errors are dominant in the available CSI, then the deterministic norm bounded error model should be considered. In this case, the optimisation problem to maximise the EE maximisation problem is formulated as

$$\max_{\mathbf{V}_{\mathrm{RF}},\mathbf{V}_{\mathrm{BB}},\mathbf{W}_{\mathrm{RF}},\mathbf{W}_{\mathrm{BB}}} \min_{\Delta} \ \mathrm{EE}(\mathbf{V}_{\mathrm{RF}},\mathbf{V}_{\mathrm{BB}},\mathbf{W}_{\mathrm{RF}},\mathbf{W}_{\mathrm{BB}}) \tag{7.1a}$$

$$\text{s.t.} \mathbf{W}_{\mathrm{RF}} \in \mathcal{W}_{\mathrm{RF}} \tag{7.1b}$$

$$\mathbf{V}_{\mathrm{RF}} \in \mathcal{V}_{\mathrm{RF}} \tag{7.1c}$$

$$\|\mathbf{V}_{j,\mathrm{RF}}\mathbf{V}_{j,\mathrm{BB}}\|_F^2 \leq P_j, \ \forall\, j \tag{7.1d}$$

$$\{\Delta_{k,j} : \|\Delta_{k,j}\|_F \leq \varepsilon_{k,j}\}, \ \forall k,j, \tag{7.1e}$$

where δ represents the error in channel measurements and ε denotes the upper limit on the Frobenius norm of the channel error. Solving the problem described by equations (7.1a)–(7.1d) is not trivial as it involves solving a multi-layered max–min problem when $\mathrm{EE}(\mathbf{V}_{\mathrm{RF}},\mathbf{V}_{\mathrm{BB}},\mathbf{W}_{\mathrm{RF}},\mathbf{W}_{\mathrm{BB}})$ is expanded as in Lemma 4.1. The analytical difficulties involved in solving the above problem is an important research direction for practical mmWave beamforming designs.

- The RS study in Chapter 5 designs private and common hybrid precoders

applicable to imperfect CSI scenarios of the mmWave system and shows that the RS transmission strategy can result in spectral efficiency gains over conventional linear transmission schemes in mmWave communications. However, it does not provide a complete set of expressions to quantify the achievable gains. Therefore, a theoretical study that provides derivations which quantify the rate and DoF gains will be complementary to the results presented in Chapter 5.

- The study of chapter 4 utilises the stochastic error model to describe the imperfect channel scenario while the study of 5 relies on the compressed sensing techniques for the estimated channel. However, it is also possible to develop enhanced channel estimation schemes. Work in this direction can focus on applying beamspace multiple signal classification method to achieve high resolution estimates of the angles of arrival and departure or exploiting the mmWave channel sparsity and developing codebook designs.

7.2.4 Interference Cancellation Techniques

- The study in Chapter 5 proposes the RS transmission as the strategy for application in multi-user mmWave systems transmissions with imperfect CSI scenarios to boost the achievable spectral efficiency. However, it relies on successive interference cancellation at the receiver which is sensitive in terms of the processing complexity and hardware requirements which scale with the number of antennas. Therefore, it will be of interest to extend the work of Chapter 5 by investigating less complex interference cancellation techniques.

7.2.5 Quality of service considerations

The study in Chapter 4 focused on the EE maximisation hybrid transceiver design and highlighted how efficiently the mmWave spectrum can be utilised. Further work which considers the issue of quality of service will be of great interest as it will ensure that each user is guaranteed their desired rate with minimum transmit power. This can be achieved by solving a sum-power minimisation problem.

For the system model considered in Chapter 4, the sum-power minimization

problem is formulated as

$$\min_{\mathbf{F}} \sum\nolimits_{j=1}^{K} \mathrm{tr}\{\mathbf{F}_j\mathbf{F}_j^H\} \tag{7.2a}$$

$$\text{s.t.}\mathcal{I}_j(\mathbf{F}) \geq R_j,\ \forall\, j, \tag{7.2b}$$

where R_j is the desired rate at receive device j and $\mathcal{I}_j$ denotes the corresponding achievable rate given as

$$\mathcal{I}_j = \log\left|\mathbf{I}_{N_\mathrm{R}} + \mathbf{\Sigma}_j^{-1}\mathbf{H}_{j,j}\mathbf{F}_j\mathbf{F}_j^H\mathbf{H}_{j,j}^H\right|. \tag{7.3}$$

Similar to the EE maximization problem, we can establish a relationship between the WSR and WMMSE, and reformulate the sum-power minimization problem under individual MSE weights as follows

$$\min_{\mathbf{F},\mathbf{W},\mathbf{Q}} \sum\nolimits_{j=1}^{K} \mathrm{tr}\{\mathbf{F}_j\mathbf{F}_j^H\} \tag{7.4a}$$

$$\text{s.t.}\quad \mathrm{tr}\left\{\mathbf{Q}_j\mathbf{M}_j\right\} - \log\left|\ln 2\,\mathbf{Q}_j\right| - \frac{N_\mathrm{s}}{\ln 2} \leq -R_j,\ \forall\, j, \tag{7.4b}$$

where $\mathbf{Q}_j$ represents the weight matrix associated with device j, and $\mathbf{M}_j$ denotes the MSE matrix of the jth transmitter–receiver pair as defined in equation (4.36). Solving the sum-power minimisation problem will be worthwhile and complementary to the results presented in Chapter 4.

7.2.6 Hardware impairment characterisation

- The work on hardware impairments in Chapter 6 offers a practical standpoint for the design of hybrid beamformers for mmWAve systems. However, our analysis is restricted to a single-user MIMO configuration, giving fundamental insights into the effect of the inevitable transceiver hardware impairments. Consequently, it will be of interest to extend the work of this chapter to a more general multi-user scenario for a wider application. Considering K users with the transceiver impairments in equation (6.12), we can adapt it as

$$\mathcal{R} = \sum_{k=1}^{K} \log_2\left(\left|\mathbf{I}_{N_\mathrm{s}} + \frac{\rho}{\sigma^2 N_\mathrm{s}}\mathbf{R}_k^{-1}\mathbf{W}_{\Omega_k}^H\mathbf{H}_k\mathbf{V}_{\Psi_k}\mathbf{V}_{\Psi_k}^H\mathbf{H}_k^H\mathbf{W}_{\Omega_k}\right|\right), \tag{7.5}$$

where $\mathbf{R}_k$ is the interference plus noise covariance matrix, given as

$$\mathbf{R}_k = \sum_{i=1, i\neq k}^{K} \mathbf{W}_{\Omega_k}^H \mathbf{H}_k \mathbf{V}_{\Psi_i} \mathbf{V}_{\Psi_i}^H \mathbf{H}_k^H \mathbf{W}_{\Omega_k} + \left(\kappa_{\mathrm{T}}^2 \frac{\rho K}{N_{\mathrm{s}}} + \kappa_{k,\mathrm{R}}^2 \rho \mathbf{V}_{\Psi_k} \mathbf{V}_{\Psi_k}^H \right) \times \mathbf{W}_{\Omega_k}^H \mathbf{H}_k \mathbf{H}_k^H \mathbf{W}_{\Omega_k} + \xi_z \mathbf{W}_{\Omega_k}^H \mathbf{W}_{\Omega_k}, \quad (7.6)$$

and the remaining variables are defined as in equation (6.12). To design the hybrid beamformers that maximise the sum rate in equation (7.5) is not trivial, as the generalised interference-cancellation methods such as zero-forcing will not sufficiently deal with the hardware impairments from other users. Hence, this additional interference source needs to be characterised with appropriate models. Investigations into this modelling and the resulting optimisation problem will be a welcome extension to the work of Chapter 6.

- For the rate-maximisation problem in equations (6.16a)–(6.16c), we provided an algorithm with a relaxed condition that assumes $\mathbf{V}_\Psi \mathbf{V}_\Psi^H \approx \mathbf{I}$. This approximation is valid for our approach of minimising the Euclidean distance between the optimal unconstrained precoder and the hybrid precoder $\mathbf{V}_\Psi$. Further work on more generalised hybrid precoder designs has the potential to further improve the results of Chapter 6 and provide more insights. In such cases, we need to solve

$$\max_{\mathbf{V}_{\mathrm{RF}}^{\mathrm{opt}}, \mathbf{V}_{\mathrm{BB}}^{\mathrm{opt}}} \log_2 \left(\left| \mathbf{I}_{N_{\mathrm{s}}} + \frac{\rho}{\sigma^2 N_{\mathrm{s}}} \frac{\mathbf{H} \mathbf{V}_\Psi \mathbf{V}_\Psi^H \mathbf{H}^H}{\left(\kappa_{\mathrm{T}}^2 \frac{\rho}{N_{\mathrm{s}}} + \kappa_{\mathrm{R}}^2 \rho \mathbf{V}_\Psi \mathbf{V}_\Psi^H \right) \mathbf{H}\mathbf{H}^H} \right| \right) \quad (7.7a)$$

$$\text{s.t.} \quad \mathbf{V}_{\mathrm{RF}} \in \mathcal{V}_{\mathrm{RF}} \quad (7.7b)$$

$$\|\mathbf{V}_\Psi\|_F^2 = N_{\mathrm{s}}, \quad (7.7c)$$

where $\mathbf{V}_\Psi = \mathbf{V}_{\mathrm{RF}} \boldsymbol{\Psi} \mathbf{V}_{\mathrm{BB}}$, which is non-trivial since $\mathbf{V}_\Psi$ appears both in the numerator and denominator of the objective function and has a non-convex constraint.

List of Publications

- O. Y. Kolawole, S. Biswas, K. Singh and T. Ratnarajah, "Transceiver design for energy-efficiency maximization in mmWave MIMO IoT Networks under minor revision, *IEEE Transactions on Green Communications and Networking*, Jan. 2019.

- O. Y. Kolawole, S. Vuppala and T. Ratnarajah, "Multiuser Millimeter Wave Cloud Radio Access Networks With Hybrid Precoding," *IEEE Systems Journal*, vol. 12, no. 4, pp. 3661-3672, Dec. 2018.

- O. Y. Kolawole, A. Papazafeiropoulos and T. Ratnarajah, "A rate-splitting strategy for multi-user millimeter wave systems with imperfect CSI", In Proc. *IEEE 19th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, June, 2018.

- O. Y. Kolawole, A. Papazafeiropoulos and T. Ratnarajah, "Impact of hardware impairments on mmWave MIMO systems with hybrid precoding", In Proc. *IEEE Wireless Communications and Networking Conference (WCNC)*, April, 2018.

- O. Y. Kolawole, S. Vuppala, M. Sellathurai and T. Ratnarajah, "On the Performance of Cognitive Satellite-Terrestrial Networks," *IEEE Transactions on Cognitive Communications and Networking*, vol. 3, no. 4, pp. 668-683, Dec. 2017.[1]

[1]This work is not included in the thesis.

Bibliography

[1] "Cisco visual networking index: Global mobile data traffic forecast update, 2017-2022 white paper," 2019. [Online]. Available: http://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/white-paper-c11-738429.pdf

[2] "Ericsson mobility report," 2018. [Online]. Available: https://www.ericsson.com/assets/local/mobility-report/documents/2018/ericsson-mobility-report-june-2018.pdf

[3] W. Xiang, K. Zheng, and X. Shen, *5G mobile communications*. Springer, 2016.

[4] T. S. Rappaport, R. W. Heath Jr, R. C. Daniels, and J. N. Murdock, *Millimeter Wave Wireless Communications*. Pearson Education, 2014.

[5] A. Goldsmith, S. A. Jafar, I. Maric, and S. Srinivasa, "Breaking spectrum gridlock with cognitive radios: An information theoretic perspective," *Proc. IEEE*, vol. 97, no. 5, pp. 894–914, May 2009.

[6] T. Rappaport, *Wireless Communications: Principles and Practice*. 2nd ed. Englewood Cliffs, NJ, USA: Prentice Hall, 2002.

[7] R. W. Heath, N. Gonzlez-Prelcic, S. Rangan, W. Roh, and A. M. Sayeed, "An overview of signal process. techniques for millimeter wave MIMO systems," *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 3, pp. 436–453, Apr 2016.

[8] M. Cudak, T. Kovarik, T. A. Thomas, A. Ghosh, Y. Kishiyama, and T. Nakamura, "Experimental mmwave 5g cellular system," in *IEEE Globecom Workshops (GC Wkshps)*, Dec 2014, pp. 377–381.

[9] G. M. Rebeiz, S. Kim, O. Inac, W. Shin, O. Gurbuz, Y. Ou, F. Golcuk, T. Kanar, and B. Ku, "Millimeter-wave large-scale phased-arrays for 5G systems," in *IEEE MTT-S International Microwave Symposium*, May 2015, pp. 1–3.

[10] J. Singh and S. Ramakrishna, "On the feasibility of beamforming in millimeter wave communication systems with multiple antenna arrays," in *IEEE Global Communications Conference*, Dec 2014, pp. 3802–3808.

[11] T. Bai, A. Alkhateeb, and R. W. Heath, "Coverage and capacity of millimeter-wave cellular networks," *IEEE Commun. Mag.*, vol. 52, no. 9, pp. 70–77, Sep 2014.

[12] M. R. Akdeniz, Y. Liu, M. K. Samimi, S. Sun, S. Rangan, T. S. Rappaport, and E. Erkip, "Millimeter wave channel modeling and cellular capacity evaluation," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1164–1179, Jun 2014.

[13] T. S. Rappaport, S. Sun, R. Mayzus, H. Zhao, Y. Azar, K. Wang, G. N. Wong, J. K. Schulz, M. Samimi, and F. Gutierrez, "Millimeter wave mobile communications for 5G cellular: It will work!" *IEEE Access*, vol. 1, pp. 335–349, May 2013.

[14] S. Rangan, T. S. Rappaport, and E. Erkip, "Millimeter-wave cellular wireless networks: Potentials and challenges," *Proc. IEEE*, vol. 102, no. 3, pp. 366–385, Mar 2014.

[15] Y. Niu, Y. Li, D. Jin, L. Su, and A. V. Vasilakos, "A survey of millimeter wave communications (mmwave) for 5G: opportunities and challenges," *Wireless Networks*, vol. 21, no. 8, pp. 2657–2676, 2015.

[16] A. Gudipati, D. Perry, L. E. Li, and S. Katti, "SoftRAN: Software defined radio access network," in *Proceedings of the second ACM SIGCOMM workshop on Hot topics in software defined networking (HotSDN)*, 2013, pp. 25–30.

[17] "Wireless HD specification overview tech. rep." 2010. [Online]. Available: http://www.wirelesshd.org/pdfs/WirelessHD-Specification-Overview-v1.1May2010.pdf

[18] S. Singh, M. N. Kulkarni, A. Ghosh, and J. G. Andrews, "Tractable model for rate in self-backhauled millimeter wave cellular networks," *IEEE J. Sel. Areas Commun.*, vol. 33, no. 10, pp. 2196–2211, Oct 2015.

[19] S. Hur, T. Kim, D. J. Love, J. V. Krogmeier, T. A. Thomas, and A. Ghosh, "Millimeter wave beamforming for wireless backhaul and access in small cell networks," *IEEE Trans. Commun.*, vol. 61, no. 10, pp. 4391–4403, Oct 2013.

[20] Z. Pi and F. Khan, "An introduction to millimeter-wave mobile broadband systems," *IEEE Commun. Mag.*, vol. 49, no. 6, pp. 101–107, Jun 2011.

[21] W. Roh, J. Y. Seol, J. Park, B. Lee, J. Lee, Y. Kim, J. Cho, K. Cheun, and F. Aryanfar, "Millimeter-wave beamforming as an enabling technology for 5G cellular communications: theoretical feasibility and prototype results," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 106–113, Feb 2014.

[22] T. Bai and R. W. Heath, "Coverage and rate analysis for millimeter-wave cellular networks," *IEEE Trans. Wireless Commun.*, vol. 14, no. 2, pp. 1100–1114, Feb 2015.

[23] E. Torkildson, C. Sheldon, U. Madhow, and M. Rodwell, "Millimeter-wave spatial multiplexing in an indoor environment," in *IEEE Globecom Workshops (GC Wkshps.)*, Nov 2009, pp. 1–6.

[24] M. Gerla, E. Lee, G. Pau, and U. Lee, "Internet of vehicles: From intelligent grid to autonomous cars and vehicular clouds," in *IEEE World Forum on Internet of Things (WF-IoT)*, Mar 2014, pp. 241–246.

[25] P. Kumari, N. Gonzalez-Prelcic, and R. W. Heath, "Investigating the IEEE 802.11ad standard for millimeter wave automotive radar," in *IEEE 82nd Vehicular Technology Conference*, Sep 2015, pp. 1–5.

[26] K. Venugopal, M. C. Valenti, and R. W. Heath, "Device-to-device millimeter wave communications: Interference, coverage, rate, and finite topologies," *IEEE Trans. Wireless Commun.*, vol. 15, no. 9, pp. 6175–6188, Sep 2016.

[27] S. Chen, H. Xu, D. Liu, B. Hu, and H. Wang, "A vision of IoT: Applications, challenges, and opportunities with china perspective," *IEEE Internet Things J.*, vol. 1, no. 4, pp. 349–359, Aug 2014.

[28] O. E. Ayach, S. Rajagopal, S. Abu-Surra, Z. Pi, and R. W. Heath, "Spatially sparse precoding in millimeter wave MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 13, no. 3, pp. 1499–1513, Mar 2014.

[29] P. Smulders and L. Correia, "Characterisation of propagation in 60 GHz radio channels," *Electronics & communication engineering journal*, vol. 9, no. 2, pp. 73–80, 1997.

[30] H. Xu, V. Kukshya, and T. S. Rappaport, "Spatial and temporal characteristics of 60-GHz indoor channels," *IEEE J. Sel. Areas Commun.*, vol. 20, no. 3, pp. 620–630, Apr 2002.

[31] A. Sayeed and J. Brady, "Beamspace MIMO for high-dimensional multiuser communication at millimeter-wave frequencies," in *IEEE Global Communications Conference*, Dec 2013, pp. 3679–3684.

[32] J. Mo, P. Schniter, N. G. Prelcic, and R. W. Heath, "Channel estimation in millimeter wave mimo systems with one-bit quantization," in *48th Asilomar Conference on Signals, Systems and Computers*, Nov 2014, pp. 957–961.

[33] T. S. Rappaport, G. R. MacCartney, M. K. Samimi, and S. Sun, "Wideband millimeter-wave propagation measurements and channel models for future wireless communication system design," *IEEE Trans. Commun.*, vol. 63, no. 9, pp. 3029–3056, Sep 2015.

[34] P. Schniter and A. Sayeed, "Channel estimation and precoder design for millimeter-wave communications: The sparse way," in *48th Asilomar Conference on Signals, Systems and Computers*, Nov 2014, pp. 273–277.

[35] T. Bai, R. Vaze, and R. W. Heath, "Analysis of blockage effects on urban cellular networks," *IEEE Trans. Wireless Commun.*, vol. 13, no. 9, pp. 5070–5083, Sep 2014.

[36] 3GPP, "Further advancements for E-UTRA physical layer aspects," TR 36.814 (release 9), 2010.

[37] M. B. Zid, K. Raoof, and A. Bouallègue, "MIMO spectral efficiency over energy consumption requirements: Application to WSNs," *International J. Commun. Netw. Syst. Sciences*, vol. 5, no. 2, pp. 121–129, 2012.

[38] Y. Li, B. Bakkaloglu, and C. Chakrabarti, "A comprehensive energy model and energy-quality evaluation of wireless transceiver front-ends," in *IEEE Workshop on Signal Processing Systems Design and Implementation*, Nov 2005, pp. 262–267.

[39] F. Rivet, Y. Deval, J. Begueret, D. Dallet, P. Cathelin, and D. Belot, "The experimental demonstration of a SASP-based full software radio receiver," *IEEE J. of Solid-State Circuits*, vol. 45, no. 5, pp. 979–988, May 2010.

[40] D. C. Daly, P. P. Mercier, M. Bhardwaj, A. L. Stone, Z. N. Aldworth, T. L. Daniel, J. Voldman, J. G. Hildebrand, and A. P. Chandrakasan, "A pulsed UWB receiver SoC for insect motion control," *IEEE J. Solid-State Circuits*, vol. 45, no. 1, pp. 153–166, Jan 2010.

[41] "IEEE standard for information technology–telecommunications and information exchange between systems–local and metropolitan area networks–specific requirements-part 11: Wireless lan medium access control (MAC) and physical layer (PHY) specifications amendment 3: Enhancements for very high throughput in the 60 ghz band," *IEEE Std 802.11ad-2012 (Amendment to IEEE Std 802.11-2012, as amended by IEEE Std 802.11ae-2012 and IEEE Std 802.11aa-2012)*, pp. 1–628, Dec 2012.

[42] M. N. Kulkarni, A. Ghosh, and J. G. Andrews, "A comparison of MIMO techniques in downlink millimeter wave cellular networks with hybrid beamforming," *IEEE Trans. Commun.*, vol. 64, no. 5, pp. 1952–1967, May 2016.

[43] S. Han, C. l. I, Z. Xu, and C. Rowell, "Large-scale antenna systems with hybrid analog and digital beamforming for millimeter wave 5G," *IEEE Commun. Mag.*, vol. 53, no. 1, pp. 186–194, Jan 2015.

[44] A. Alkhateeb, O. E. Ayach, G. Leus, and R. W. Heath, "Hybrid precoding for millimeter wave cellular systems with partial channel knowledge," in *2013 Information Theory and Applications Workshop (ITA)*, Feb 2013, pp. 1–5.

[45] S. Reyaz, C. Samuelsson, R. Malmqvist, M. Kaynak, and A. Rydberg, "Millimeter-wave RF-MEMS SPDT switch networks in a SiGe BiCMOS process technology," in *2012 7th European Microwave Integrated Circuit Conference*, Oct 2012, pp. 691–694.

[46] R. Mndez-Rial, C. Rusu, A. Alkhateeb, N. Gonzlez-Prelcic, and R. W. Heath, "Channel estimation and hybrid combining for mmwave: Phase shifters or switches?" in *Information Theory and Applications Workshop (ITA)*, Feb 2015, pp. 90–97.

[47] X. Gao, O. Edfors, F. Tufvesson, and E. G. Larsson, "Multi-switch for antenna selection in massive mimo," in *2015 IEEE Global Communications Conference*, Dec 2015, pp. 1–6.

[48] A. Sayeed and N. Behdad, "Continuous aperture phased MIMO: Basic theory and applications," in *48th Annual Allerton Conference on Communication, Control, and Computing*, Sep 2010, pp. 1196–1203.

[49] J. Brady, N. Behdad, and A. M. Sayeed, "Beamspace MIMO for millimeter-wave communications: System architecture, modeling, analysis, and measurements," *IEEE Trans. Antennas Propag.*, vol. 61, no. 7, pp. 3814–3827, Jul 2013.

[50] W. Hong, Z. H. Jiang, C. Yu, J. Zhou, P. Chen, Z. Yu, H. Zhang, B. Yang, X. Pang, M. Jiang, Y. Cheng, M. K. T. Al-Nuaimi, Y. Zhang, J. Chen, and S. He, "Multibeam antenna technologies for 5G wireless communications," *IEEE Trans. Antennas Propag.*, vol. 65, no. 12, pp. 6231–6249, Dec 2017.

[51] M. A. B. Abbasi, V. F. Fusco, H. Tataria, and M. Matthaiou, "Constant-r lens beamformer for low-complexity millimeter-wave hybrid MIMO," *IEEE Trans. Microw. Theory Techn.*, pp. 1–10, 2019.

[52] J. Mo and R. W. Heath, "Capacity analysis of one-bit quantized MIMO systems with transmitter channel state information," *IEEE Trans. Signal Process.*, vol. 63, no. 20, pp. 5498–5512, Oct 2015.

[53] J. Singh, O. Dabeer, and U. Madhow, "On the limits of communication with low-precision analog-to-digital conversion at the receiver," *IEEE Trans. Commun.*, vol. 57, no. 12, pp. 3629–3639, Dec 2009.

[54] Y. Li, C. Tao, G. Seco-Granados, A. Mezghani, A. L. Swindlehurst, and L. Liu, "Channel estimation and performance analysis of one-bit massive MIMO systems," *IEEE Trans. Signal Process.*, vol. 65, no. 15, pp. 4075–4089, Aug 2017.

[55] A. Mezghani and J. A. Nossek, "On ultra-wideband MIMO systems with 1-bit quantized outputs: Performance analysis and input optimization," in *IEEE International Symposium on Information Theory*, Jun 2007, pp. 1286–1289.

[56] D. Verenzuela, E. Bjrnson, and M. Matthaiou, "Hardware design and optimal adc resolution for uplink massive MIMO systems," in *IEEE Sensor Array and Multichannel Signal Processing Workshop (SAM)*, July 2016, pp. 1–5.

[57] A. Alkhateeb, J. Mo, N. Gonzalez-Prelcic, and R. W. Heath, "MIMO precoding and combining solutions for millimeter-wave systems," *IEEE Commun. Mag.*, vol. 52, no. 12, pp. 122–131, Dec 2014.

[58] J. Wang, Z. Lan, C. woo Pyo, T. Baykas, C. sean Sum, M. A. Rahman, J. Gao, R. Funada, F. Kojima, H. Harada, and S. Kato, "Beam codebook based beamforming protocol for multi-gbps millimeter-wave WPAN systems," *IEEE J. Sel. Areas Commun.*, vol. 27, no. 8, pp. 1390–1399, Oct 2009.

[59] Y. M. Tsang, A. S. Y. Poon, and S. Addepalli, "Coding the beams: Improving beamforming training in mmwave communication system," in *IEEE Global Telecommunications Conference*, Dec 2011, pp. 1–6.

[60] A. Alkhateeb, O. E. Ayach, G. Leus, and R. W. Heath, "Channel estimation and hybrid precoding for millimeter wave cellular systems," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 831–846, Oct 2014.

[61] D. Tse and P. Viswanath, *Fundamentals of wireless communication*. Cambridge university press, 2005.

[62] M. L. Malloy and R. D. Nowak, "Near-optimal adaptive compressed sensing," *IEEE Trans. Inf. Theory*, vol. 60, no. 7, pp. 4001–4012, Jul 2014.

[63] A. Mezghani, F. Antreich, and J. A. Nossek, "Multiple parameter estimation with quantized channel output," in *International ITG Workshop on Smart Antennas (WSA)*, Feb 2010, pp. 143–150.

[64] O. Y. Kolawole, S. Vuppala, and T. Ratnarajah, "Multiuser millimeter wave cloud radio access networks with hybrid precoding," *IEEE Syst. J.*, vol. 12, no. 4, pp. 3661–3672, Dec 2018.

[65] C. Sacchi, T. F. Rahman, N. Bartolomei, S. Morosi, A. Mazzinghi, and F. Ciabini, "Design and assessment of a CE-OFDM-based mm-wave 5G communication system," in *IEEE Globecom Workshops (GC Wkshps)*, Dec 2016, pp. 1–7.

[66] A. S. Cacciapuoti, R. Subramanian, K. R. Chowdhury, and M. Caleffi, "Software-defined network controlled switching between millimeter wave and terahertz small cells," *arXiv preprint arXiv:1702.02775*, 2017.

[67] T. S. Rappaport, F. Gutierrez, E. Ben-Dor, J. N. Murdock, Y. Qiao, and J. I. Tamir, "Broadband millimeter-wave propagation measurements and models using adaptive-beam antennas for outdoor urban cellular communications," *IEEE Trans. Antennas Propag.*, vol. 61, no. 4, pp. 1850–1859, Apr 2013.

[68] M. D. Renzo, "Stochastic geometry modeling and analysis of multi-tier millimeter wave cellular networks," *IEEE Trans. Wireless Commun.*, vol. 14, no. 9, pp. 5038–5057, Sep 2015.

[69] M. N. Kulkarni, A. Alkhateeb, and J. G. Andrews, "A tractable model for per user rate in multiuser millimeter wave cellular networks," in *49th Asilomar Conference on Signals, Systems and Computers*, Nov 2015, pp. 328–332.

[70] F. Boccardi, R. W. Heath, A. Lozano, T. L. Marzetta, and P. Popovski, "Five disruptive technology directions for 5g," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 74–80, Feb 2014.

[71] Y. Shi, J. Zhang, and K. B. Letaief, "Group sparse beamforming for green cloud-RAN," *IEEE Trans. Wireless Commun.*, vol. 13, no. 5, pp. 2809–2823, May 2014.

[72] P. Marsch, B. Raaf, A. Szufarska, P. Mogensen, H. Guan, M. Farber, S. Redana, K. Pedersen, and T. Kolding, "Future mobile communication networks: Challenges in the design and operation," *IEEE Veh. Technol. Mag.*, vol. 7, no. 1, pp. 16–23, Mar 2012.

[73] S. R. Lee, S. H. Moon, J. S. Kim, and I. Lee, "Capacity analysis of distributed antenna systems in a composite fading channel," *IEEE Trans. Wireless Commun.*, vol. 11, no. 3, pp. 1076–1086, Mar 2012.

[74] H. Zhu, "Performance comparison between distributed antenna and microcellular systems," *IEEE J. Sel. Areas Commun.*, vol. 29, no. 6, pp. 1151–1163, Jun 2011.

[75] J. Zhang and J. G. Andrews, "Distributed antenna systems with randomness," *IEEE Trans. Wireless Commun.*, vol. 7, no. 9, pp. 3636–3646, Sep 2008.

[76] H. He, J. Xue, T. Ratnarajah, F. A. Khan, and C. B. Papadias, "Modeling and analysis of cloud radio access networks using mat'ern hard-core point processes," *IEEE Trans. Wireless Commun.*, vol. 15, no. 6, pp. 4074–4087, Jun 2016.

[77] "Common public radio interfaceCPRI: Interface specification," 2013. [Online]. Available: http://cpri.info/downloads/CPRI_v_6_0_2013-08-30.pdf

[78] "5G PPP architecture working group view on 5G architecture white paper," 2016. [Online]. Available: http://5g-ppp.eu/wp-content/uploads/2014/02/5G-PPP-5G-Architecture-WP-July-2016.pdf

[79] "5G XHaul project deliverable D2.1 requirements specification and KPIs document," 2015. [Online]. Available: http://5g-xhaul-project.eu/download/5G-XHaul_D2_1.pdf

[80] A. Alkhateeb, G. Leus, and R. W. Heath, "Limited feedback hybrid precoding for multi-user millimeter wave systems," *IEEE Trans. Wireless Commun.*, vol. 14, no. 11, pp. 6481–6494, Nov 2015.

[81] A. Alkhateeb, R. W. Heath, and G. Leus, "Achievable rates of multi-user millimeter wave systems with hybrid precoding," in *IEEE International Conference on Communication Workshop (ICCW)*, Jun 2015, pp. 1232–1237.

[82] X. Yu, J. C. Shen, J. Zhang, and K. B. Letaief, "Alternating minimization algorithms for hybrid precoding in millimeter wave MIMO systems," *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 3, pp. 485–500, Apr 2016.

[83] O. E. Ayach, R. W. Heath, S. Abu-Surra, S. Rajagopal, and Z. Pi, "The capacity optimality of beam steering in large millimeter wave MIMO sys-

tems," in *IEEE International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, Jun 2012, pp. 100–104.

[84] A. Ghosh, T. A. Thomas, M. C. Cudak, R. Ratasuk, P. Moorut, F. W. Vook, T. S. Rappaport, G. R. MacCartney, S. Sun, and S. Nie, "Millimeter-wave enhanced local area systems: A high-data-rate approach for future wireless networks," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1152–1163, Jun 2014.

[85] L. Kleinrock, *Queuing Systems*. Wiley, 1975.

[86] T. Han and N. Ansari, "Network utility aware traffic load balancing in backhaul-constrained cache-enabled small cell networks with hybrid power supplies," *IEEE Trans. Mobile Comput.*, vol. 16, no. 10, pp. 2819–2832, Oct 2017.

[87] J. Wu, Z. Zhang, Y. Hong, and Y. Wen, "Cloud radio access network (C-RAN): A primer," *IEEE Network*, vol. 29, no. 1, pp. 35–41, Jan 2015.

[88] J. Bartelt, P. Rost, D. Wubben, J. Lessmann, B. Melis, and G. Fettweis, "Fronthaul and backhaul requirements of flexibly centralized radio access networks," *IEEE Wireless Commun.*, vol. 22, no. 5, pp. 105–111, Oct 2015.

[89] D. B. Cheikh, J. M. Kelif, M. Coupechoux, and P. Godlewski, "Multicellular zero forcing precoding performance in rayleigh and shadow fading," in *IEEE 73rd Vehicular Technology Conference*, May 2011, pp. 1–5.

[90] N. L. Johnson, S. Kotz, and N. Balakrishnan, "Continuous univariate distributions," *Journal of the Royal Statistical Society-Series A Statistics in Society*, vol. 159, no. 2, p. 343, 1996.

[91] A. M. Ibrahim, T. ElBatt, and A. El-Keyi, "Coverage probability analysis for wireless networks using repulsive point processes," in *IEEE 24th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)*, Sept 2013, pp. 1002–1007.

[92] M. Haenggi, *Stochastic Geometry for Wireless Networks*. Cambridge University Press, 2012.

[93] M. D. Renzo, A. Guidotti, and G. E. Corazza, "Average rate of downlink heterogeneous cellular networks over generalized fading channels: A

stochastic geometry approach," *IEEE Trans. Commun.*, vol. 61, no. 7, pp. 3050–3071, Jul 2013.

[94] J. Bartelt, D. Wübben, P. Rost, J. Lessmann, and G. Fettweis, "Fronthaul for a flexible centralization in cloud radio access networks," *Backhauling/Fronthauling for Future Wireless Systems*, pp. 55–84, 2016.

[95] S. Chen and J. Zhao, "The requirements, challenges, and technologies for 5G of terrestrial mobile telecommunication," *IEEE Commun. Mag.*, vol. 52, no. 5, pp. 36–43, May 2014.

[96] C.-W. Tsai, C.-F. Lai, and A. V. Vasilakos, "Future internet of things: open issues and challenges," *Wireless Networks*, vol. 20, no. 8, pp. 2201–2217, 2014.

[97] Y. Y. Lee, C. H. Wang, and Y. H. Huang, "A hybrid RF/baseband precoding processor based on parallel-index-selection matrix-inversion-bypass simultaneous orthogonal matching pursuit for millimeter wave MIMO systems," *IEEE Trans. Signal Process.*, vol. 63, no. 2, pp. 305–317, Jan 2015.

[98] C. E. Chen, "An iterative hybrid transceiver design algorithm for millimeter wave MIMO systems," *IEEE Wireless Commun. Lett.*, vol. 4, no. 3, pp. 285–288, Jun 2015.

[99] X. Gao, L. Dai, S. Han, C. L. I, and R. W. Heath, "Energy-efficient hybrid analog and digital precoding for mmwave mimo systems with large antenna arrays," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 4, pp. 998–1009, Apr 2016.

[100] J. Qi and S. Aissa, "Analysis and compensation of power amplifier nonlinearity in MIMO transmit diversity systems," *IEEE Trans. Veh. Technol.*, vol. 59, no. 6, pp. 2921–2931, Jul 2010.

[101] C. Studer, M. Wenk, and A. Burg, "MIMO transmission with residual transmit-RF impairments," in *International ITG Workshop on Smart Antennas (WSA)*, Feb 2010, pp. 189–196.

[102] O. Kolawole, A. Papazafeiropoulos, and T. Ratnarajah, "Impact of hardware impairments on mmwave MIMO systems with hybrid precoding," in *2018 IEEE Wireless Communications and Networking Conference (WCNC)*, Apr 2018, pp. 1–6.

[103] E. Bjornson, M. Matthaiou, and M. Debbah, "A new look at dual-hop relaying: Performance limits with hardware impairments," *IEEE Trans. Commun.*, vol. 61, no. 11, pp. 4512–4525, November 2013.

[104] E. Bjrnson, M. Matthaiou, and M. Debbah, "Massive MIMO with non-ideal arbitrary arrays: Hardware scaling laws and circuit-aware design," *IEEE Trans. Wireless Commun.*, vol. 14, no. 8, pp. 4353–4368, Aug 2015.

[105] X. Zhang, M. Matthaiou, M. Coldrey, and E. Bjrnson, "Impact of residual transmit RF impairments on training-based MIMO systems," *IEEE Trans. Commun.*, vol. 63, no. 8, pp. 2899–2911, Aug 2015.

[106] C. Masouros, M. Sellathurai, and T. Ratnarajah, "Computationally efficient vector perturbation precoding using thresholded optimization," *IEEE Trans. Commun.*, vol. 61, no. 5, pp. 1880–1890, May 2013.

[107] D. Ha, K. Lee, and J. Kang, "Energy efficiency analysis with circuit power consumption in massive mimo systems," in *IEEE 24th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)*, Sep. 2013, pp. 938–942.

[108] J. Xu, L. Qiu, and C. Yu, "Improving energy efficiency through multimode transmission in the downlink MIMO systems," *EURASIP J. Wireless Commun. Netw.*, vol. 2011, no. 1, p. 200, 2011.

[109] C. Balanis, *Antenna Theory: Analysis and Design*. Hoboken,NJ, USA: Wiley, 2012.

[110] J. Crouzeix and J. Ferland, "Algorithms for generalized fractional programming," *Mathematical Programming*, vol. 52, no. 1-3, pp. 191–207, 1991.

[111] Q. Shi, M. Razaviyayn, Z. Q. Luo, and C. He, "An iteratively weighted MMSE approach to distributed sum-utility maximization for a MIMO interfering broadcast channel," *IEEE Trans. Signal Process.*, vol. 59, no. 9, pp. 4331–4340, Sep 2011.

[112] J. Shin and J. Moon, "Weighted-sum-rate-maximizing linear transceiver filters for the K-user MIMO interference channel," *IEEE Trans.Commun.*, vol. 60, no. 10, pp. 2776–2783, Oct 2012.

[113] P. Tseng, "Dual coordinate ascent methods for non-strictly convex minimization," *Mathematical programming*, vol. 59, no. 1-3, pp. 231–247, 1993.

[114] R. Hunger, "Floating point operations in matrix-vector calculus," Tech. Rep., 2007.

[115] M. Kim and Y. H. Lee, "MSE-based hybrid RF/baseband processing for millimeter-wave communication systems in MIMO interference channels," *IEEE Trans. Veh. Technol.*, vol. 64, no. 6, pp. 2714–2720, Jun 2015.

[116] J. A. Tropp and A. C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Trans. Inf. Theory*, vol. 53, no. 12, pp. 4655–4666, Dec 2007.

[117] L. Rebollo-Neira and D. Lowe, "Optimized orthogonal matching pursuit approach," *IEEE Signal Process. Lett.*, vol. 9, no. 4, pp. 137–140, Apr 2002.

[118] J. Maurer, J. Jalden, D. Seethaler, and G. Matz, "Vector perturbation precoding revisited," *IEEE Trans. Signal Process.*, vol. 59, no. 1, pp. 315–328, Jan 2011.

[119] B. Hassibi and B. M. Hochwald, "How much training is needed in multiple-antenna wireless links?" *IEEE Trans. Inf. Theory*, vol. 49, no. 4, pp. 951–963, Apr 2003.

[120] B. P. Day, A. R. Margetts, D. W. Bliss, and P. Schniter, "Full-duplex bidirectional MIMO: Achievable rates under limited dynamic range," *IEEE Trans. Signal Process.*, vol. 60, no. 7, pp. 3702–3713, July 2012.

[121] E. Antonio-Rodrguez, R. Lpez-Valcarce, T. Riihonen, S. Werner, and R. Wichman, "SINR optimization in wideband full-duplex MIMO relays under limited dynamic range," in *IEEE 8th Sensor Array and Multichannel Signal Processing Workshop (SAM)*, June 2014, pp. 177–180.

[122] Y. Mao, B. Clerckx, and V. O. K. Li, "Energy efficiency of rate-splitting multiple access, and performance benefits over SDMA and NOMA," in *15th International Symposium on Wireless Communication Systems (ISWCS)*, Aug 2018, pp. 1–5.

[123] S. R. Sabuj and M. Hamamura, "Energy efficiency analysis of cognitive radio network using stochastic geometry," in *IEEE Conference on Standards for Communications and Networking (CSCN)*, Oct 2015.

[124] H. Q. Ngo, E. G. Larsson, and T. L. Marzetta, "Energy and spectral efficiency of very large multiuser MIMO systems," *IEEE Trans. Commun.*, vol. 61, no. 4, pp. 1436–1449, April 2013.

[125] D. Nguyen, L. Tran, P. Pirinen, and M. Latva-aho, "Precoding for full duplex multiuser MIMO systems: Spectral and energy efficiency maximization," *IEEE Trans. Signal Process.*, vol. 61, no. 16, pp. 4038–4050, Aug 2013.

[126] H. Chen, G. Li, and J. Cai, "Spectralenergy efficiency tradeoff in full-duplex two-way relay networks," *IEEE Syst. J.*, vol. 12, no. 1, pp. 583–592, March 2018.

[127] O. Kolawole, A. Papazafeiropoulos, and T. Ratnarajah, "A rate-splitting strategy for multi-user millimeter-wave systems with imperfect CSI," in *2018 IEEE 19th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, Jun 2018.

[128] M. Dai and B. Clerckx, "Multiuser millimeter wave beamforming strategies with quantized and statistical csit," *IEEE Trans. Wireless Commun.*, vol. 16, no. 11, pp. 7025–7038, Nov 2017.

[129] S. Yang, M. Kobayashi, D. Gesbert, and X. Yi, "Degrees of freedom of time correlated MISO broadcast channel with delayed CSIT," *IEEE Trans. Inf. Theory*, vol. 59, no. 1, pp. 315–328, Jan 2013.

[130] H. Joudeh and B. Clerckx, "Sum-rate maximization for linearly precoded downlink multiuser MISO systems with partial CSIT: A rate-splitting approach," *IEEE Trans. Commun.*, vol. 64, no. 11, pp. 4847–4861, Nov 2016.

[131] M. Dai, B. Clerckx, D. Gesbert, and G. Caire, "A rate splitting strategy for massive MIMO with imperfect CSIT," *IEEE Trans. Wireless Commun.*, vol. 15, no. 7, pp. 4611–4624, Jul 2016.

[132] C. Hao, Y. Wu, and B. Clerckx, "Rate analysis of two-receiver MISO broadcast channel with finite rate feedback: A rate-splitting approach," *IEEE Trans. Commun.*, vol. 63, no. 9, pp. 3232–3246, Sep 2015.

[133] A. Papazafeiropoulos, B. Clerckx, and T. Ratnarajah, "Rate-splitting to mitigate residual transceiver hardware impairments in massive MIMO systems," *IEEE Trans. Veh. Technol.*, vol. 66, no. 9, pp. 8196–8211, Sep 2017.

[134] T. Schenk, *RF Imperfections in High-Rate Wireless Systems: Impact and Digital Compensation*. Springer Science & Business Media, 2008.

[135] E. Bjornson, J. Hoydis, M. Kountouris, and M. Debbah, "Massive MIMO systems with non-ideal hardware: Energy efficiency, estimation, and capacity limits," *IEEE Trans. Inf. Theory*, vol. 60, no. 11, pp. 7112–7139, Nov 2014.

[136] H. Holma and A. Toskala, *LTE for UMTS: Evolution to LTE-Advanced*. John Wiley & Sons, 2011.

[137] Z. Xiao, T. He, P. Xia, and X. Xia, "Hierarchical codebook design for beamforming training in millimeter-wave communication," *IEEE Trans. Wireless Commun.*, vol. 15, no. 5, pp. 3380–3392, May 2016.

[138] J. Hoydis, S. Brink, and M. Debbah, "Massive MIMO in the UL/DL of cellular networks: How many antennas do we need?" *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 160–171, Feb 2013.

[139] J. Jose, A. Ashikhmin, T. L. Marzetta, and S. Vishwanath, "Pilot contamination and precoding in multi-cell TDD systems," *IEEE Trans. Wireless Commun.*, vol. 10, no. 8, pp. 2640–2651, Aug 2011.

[140] J. G. Andrews, T. Bai, M. N. Kulkarni, A. Alkhateeb, A. K. Gupta, and R. W. Heath, "Modeling and analyzing millimeter wave cellular systems," *IEEE Trans. Commun.*, vol. 65, no. 1, pp. 403–430, Jan 2017.

[141] F. W. Vook, A. Ghosh, and T. A. Thomas, "MIMO and beamforming solutions for 5G technology," in *IEEE MTT-S International Microwave Symposium*, Jun 2014, pp. 1–4.

[142] X. Zhang, M. Matthaiou, E. Bjrnson, M. Coldrey, and M. Debbah, "On the MIMO capacity with residual transceiver hardware impairments," in *IEEE International Conference on Communications (ICC)*, Jun 2014, pp. 5299–5305.

[143] U. Gustavsson, C. Sanchz-Perez, T. Eriksson, F. Athley, G. Durisi, P. Landin, K. Hausmair, C. Fager, and L. Svensson, "On the impact of hardware impairments on massive MIMO," in *IEEE Globecom Workshops (GC Wkshps)*, Dec 2014, pp. 294–300.

[144] F. Athley, G. Durisi, and U. Gustavsson, "Analysis of massive MIMO with hardware impairments and different channel models," in *29th European Conference on Antennas and Propagation (EuCAP)*, May 2015, pp. 1–5.

[145] J. Zhu, D. W. K. Ng, N. Wang, R. Schober, and V. K. Bhargava, "Analysis and design of secure massive MIMO systems in the presence of hardware impairments," *IEEE Trans. Wireless Commun.*, vol. 16, no. 3, pp. 2001–2016, Mar 2017.

[146] A. K. Papazafeiropoulos, "Impact of general channel aging conditions on the downlink performance of massive MIMO," *IEEE Trans. Veh. Technol.*, vol. 66, no. 2, pp. 1428–1442, Feb 2017.

[147] A. Papazafeiropoulos and T. Ratnarajah, "Toward a realistic assessment of multiple antenna HCNs: Residual additive transceiver hardware impairments and channel aging," *IEEE Trans. Veh. Technol.*, vol. 66, no. 10, pp. 9061–9073, Oct 2017.

[148] S. Gimenez, S. Roger, P. Baracca, D. Martín-Sacristán, J. F. Monserrat, V. Braun, and H. Halbauer, "Performance evaluation of analog beamforming with hardware impairments for mmW massive MIMO communication in an urban scenario," *IEEE Sensors J.*, vol. 16, no. 10, p. 1555, 2016.

[149] Y. Zou, P. Zetterberg, U. Gustavsson, T. Svensson, A. Zaidi, T. Kadur, W. Rave, and G. Fettweis, "Impact of major RF impairments on mm-wave communications using OFDM waveforms," in *IEEE Globecom Workshops (GC Wkshps)*, Dec 2016, pp. 1–7.

[150] A. Goldsmith, S. A. Jafar, N. Jindal, and S. Vishwanath, "Capacity limits of MIMO channels," *IEEE J. Sel. Areas Commun.*, vol. 21, no. 5, pp. 684–702, Jun 2003.

[151] A. Demir, A. Mehrotra, and J. Roychowdhury, "Phase noise in oscillators: A unifying theory and numerical methods for characterisation," in *35th Design and Automation Conference*, June 1998, pp. 26–31.

[152] A. Pitarokoilis, S. K. Mohammed, and E. G. Larsson, "Uplink performance of time-reversal mrc in massive MIMO systems subject to phase noise," *IEEE Trans. Wireless Commun.*, vol. 14, no. 2, pp. 711–723, Feb 2015.

[153] D. P. Palomar, J. M. Cioffi, and M. A. Lagunas, "Joint Tx-Rx beamforming design for multicarrier MIMO channels: A unified framework for convex optimization," *IEEE Trans. Signal Process.*, vol. 51, no. 9, pp. 2381–2401, Sept 2003.

[154] Y. C. Pati, R. Rezaiifar, and P. S. Krishnaprasad, "Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition," in *27th Asilomar Conference on Signals, Systems and Computers*, Nov 1993, pp. 40–44 vol.1.

[155] W. Ni, X. Dong, and W. Lu, "Near-optimal hybrid processing for massive MIMO systems via matrix decomposition," *IEEE Trans. Signal Process.*, vol. 65, no. 15, pp. 3922–3933, Aug 2017.

[156] T. Kailath, A. Sayed, and B. Hassibi, *Linear Estimation*. Prentice Hall Upper Saddle River, NJ, 2000, vol. 1.

[157] "Common public radio interface: eCPRI interface specification," 2018. [Online]. Available: http://cpri.info/downloads/eCPRI_v_1_2_2018_06_25.pdf